Cognitive Flexibility and Persistent Pain



By

Caitlin Alyce Howlett, BPsych (Hons)
ORCID: 000-0002-4584-8641

A thesis completed and submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of South Australia

Innovation, Implementation & Clinical Translation (IIMPACT) in Health

Allied Health and Human Performance

Kaurna Country, Adelaide, South Australia, Australia

April 2023

University of South Australia

This thesis is dedicated to my parents, Kim and Richard, as well as my partner, Mitchell, for their support, sacrifice, and unconditional love.

I also dedicate this thesis to my grandfather, Peter Cropley.

Thank you for instilling in me that no dream is ever too far out of reach.
I cannot express how much I wish you were here to share this achievement with me.

♡

Table of Contents

Summary vi

Declaration Statement vii

Acknowledgements viii

List of Figures xi

List of Tables xiii

Preface xiv

CHAPTER ONE: Introduction 1

1.1 Pain 2

1.1.1 Definition 2

1.1.2 Pain and nociception 2

1.1.3 When does acute pain transition into persistent pain? 3

1.1.4 The biopsychosocial model of pain 3

1.1.5 Prevalence, burden and risk factors 6

1.2 When pain persists – chronic pain and learning 9

1.2.1 Long-term potentiation – 'neurones that fire together, wire together' 9

1.2.2 Associative learning 10

1.2.3 Predictive processing: What we expect influences what we get 11

1.3 Cognitive flexibility 12

1.3.1 Definition 12

1.3.2 This thesis is not about psychological flexibility 14

1.3.3 The lack of definitional consensus surrounding cognitive flexibility 16

1.3.4 The development and decline of cognitive flexibility across the lifespan 16

1.3.5 Cognitive flexibility and its relation with cognitive abilities, psychopathology and physical aspects of health 17

1.3.6 Assessing cognitive flexibility 17

1.3.6.1 Neuropsychological assessments 18

1.3.6.2 Self-report questionnaires 18

1.3.7 The Wisconsin Card Sorting Test – an attempt to 'set the record straight' 19

1.3.8 The relationship between, and conflation of, self-report and neuropsychological assessments of cognitive flexibility 22

1.4 Cognitive flexibility and persistent pain 25

1.4.1 Are people with persistent pain characterised by poor cognitive flexibility? 25

1.5 Thesis aim and specific research objectives 28

CHAPTER TWO: Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility 30

CHAPTER THREE: Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults 31

CHAPTER FOUR: Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis 32

CHAPTER FIVE: Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy 33

CHAPTER SIX: Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study 34

CHAPTER SEVEN: Discussion 75

7.1 Brief overview 76

7.2 Summary of the main findings 76

7.3 Implications 78

7.3.1 Reducing inconsistency and ambiguity through precise terminology, definitions and language 79

7.3.2 There is a need for increased transparency surrounding commercial and non-commercial versions of the Wisconsin Card Sorting Test 81

7.3.3 Confounders: Are key pieces of the puzzle being missed? 81

7.4 Strengths, limitations, caveats, and speculations 81

7.4.1 Could cognitive flexibility assessments be related after all? 82

7.4.2 Compliance with Open Science practices: Transparency, replicability and reproducibility 82

7.4.3 Drawbacks of various study designs 83

7.4.4 Limited options for critical appraisal tools 84

7.4.5 Inherent issues of self-report and neuropsychological assessments 85

7.5 Unforeseen challenges and associated learnings during the PhD 86

7.5.1 Benefits and barriers of statistical and experimental software 86

7.5.2 Extracting data for inclusion in systematic reviews 87

7.5.3 Impact of the COVID-19 pandemic 88

7.6 Directions for future research 88

7.6.1 Consistent terminology and a uniform definition of cognitive flexibility is required 88

7.6.2 Next on the agenda: Develop relevant and ecologically valid assessments of cognitive flexibility 89
7.6.3 Is the standardised manual of the Wisconsin Card Sorting Test in need of another revision? 91
7.7 Concluding remarks 99
7.8 References 101

CHAPTER EIGHT: Research dissemination: Publications, presentations, awards, impact and additional activities 128
8.1 Research outputs 129
8.2 Presentations during candidature 130
8.3 Scholarships and awards attained during candidature 131
8.4 Impact of studies contained in this thesis 131
8.5 Additional activities during candidature 132

CHAPTER NINE: Appendices 133
Appendix A: Supplementary material for Chapters Three and Four 134
Appendix B: Supplementary material for Chapter Three 152
Appendix C: Supplementary material for Chapter Four 168
Appendix D: Supplementary material for Chapter Six 187
Appendix E: Permissions 248
Appendix F: Statements of Authorship 267

Summary

Persistent pain is a burdensome public health problem that affects a large proportion of the population and is a leading cause of disability worldwide. It is characterised by pain that lasts for three months or more, and that no longer has protective value for bodily tissues. When pain persists, it interferes with many facets of a person's life and results in substantial economic impact at both the individual and societal level. The exact cognitive markers that predispose a person to the development and maintenance of persistent pain is not entirely understood. Further research is required to better understand the role and importance of cognitive markers so as to improve the prevention and treatment of painful conditions. *Cognitive flexibility* – a latent construct that broadly refers to the ability to effectively adapt cognitive and behavioural strategies when environmental or task demands change – is one marker that may hold promise. The broad aim of this thesis was to better understand the potential role of cognitive flexibility in persistent pain.

Chapter 1 introduces the broader context of this work, the key concepts, and relevance of studying cognitive flexibility with persistent pain in mind. Chapter 2 is a critical review that highlights the issues with the current use of a popular neuropsychological assessment of cognitive flexibility – the Wisconsin Card Sorting Test (WCST) – particularly, the implementation of different scoring procedures and the inconsistent use of key terminology across the literature. Chapters 3 and 4 each present meta-analytic reviews that evaluated the relationship between self-report and neuropsychological assessments of cognitive flexibility in non-clinical (Chapter 3) and clinical (Chapter 4) cohorts. Both studies suggest that these two assessments are unlikely to be capturing the same underlying construct of 'cognitive flexibility' and thus, cannot be considered as equivalent and interchangeable proxies for one another. Chapter 5 offers a critical historical review that seeks to understand and explain when, how and why self-report and neuropsychological assessments of cognitive flexibility have become conflated. That work identified key and unvalidated assumptions made by researchers and clinicians, mainly on the basis of similar labels being used for different things. Chapter 6 provides an online, cross-sectional, observational study that investigates whether evaluations on a self-report questionnaire and performance on a neuropsychological assessment of 'cognitive flexibility' are poorer in people with persistent pain than pain-free controls, and finds that people with persistent pain do not appear to exhibit such deficits. Chapter 7 provides a general discussion of the key findings of this thesis and offers theoretical and practical implications of this work, as well as possible future directions.

Declaration Statement

I declare that this thesis presents work carried out by myself and does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; to the best of my knowledge, it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, are clearly acknowledged.



Caitlin A. Howlett 4th April, 2023

Acknowledgements

I would like to acknowledge the Traditional Custodians of the land upon which this work was conducted, the Kaurna people of the Adelaide region, and pay my respects to their Elders past, present and emerging.

This doctoral thesis marks the completion of a gruelling, yet deeply fulfilling journey – one that has undoubtedly changed my life in ways that I never thought possible. I feel very fortunate that my academic career has been one filled with vast opportunities, challenges, learnings, and much personal and professional growth. The work contained in this thesis has taken me to places both in and outside of Australia; in April 2021, it took me to a mountainous and picturesque Tasmania, and in March 2023, it took me to the cobblestoned streets of a snowy and medieval Belgium, where on both occasions, I had the privilege of sharing my passion with like-minded people. The highest of highs, the lowest of lows and all the experiences in between have helped shape me into the scientist I am today, and for that, I am so grateful.

Like any scientific endeavour, this thesis was made possible thanks to the support and generosity of many people over the last four years. Some words of appreciation …

First and foremost, to my supervisory team. Lorimer – throughout the years you've occasionally reminded me that this PhD is not going to change the world, but what you may not know is just how much this PhD has changed *my world.* Kudos to you for fostering a research team that holds integrity, innovation, collaboration, communication and excellence at its core – being a member of this group will always remain one of my most treasured experiences, all thanks to you and the folks in the BiM orbit. Your solid commitment to make a difference*,* no matter how big or small, and to do *things that matter*, is truly inspiring. I am tremendously grateful for your expertise, compassion, guidance, and unwavering belief in me. I know, wholeheartedly, that doing a PhD was a great thing to have done, and an even better thing to have done with *you* as my mentor, which makes the closing of this chapter all the more bittersweet. Valeria – it was your initial suggestion that encouraged me to pursue a PhD in the field of pain, and without it, I may have never considered taking this path. Thank you for your ongoing encouragement, support, and friendship over the years. Ina – although this thesis has taken many unexpected twists and turns, I appreciate all that you have done for me. I will always be thankful for the role that each of you played in my academic career. I hope this work makes you all proud.

To my end-user – Tim Cocks, from the Neuro Orthopaedic Institute. You never turned down an opportunity to join me for a coffee to converse about our shared interest in predictive processing theory; thank you, those conversations have been invaluable.

To my collaborators – I'm so thankful for the hard work you invested into making each manuscript come to fruition. A special mention to a few collaborators in particular. Stephanie Miles – my sincerest thanks for your collaboration, friendship, and shared passion for improving the field; I feel lucky to have crossed paths with you on this journey, and I am hopeful that we will remain life-long colleagues. Carolyn Berryman – thank you for being someone I could turn to when I felt overwhelmed by the deep rabbit holes that I often found myself in. Michael Wewege and Ty Stanford – you've both contributed to my immense appreciation for R. I am so grateful for your generosity, kindness and patience as I navigated what was truly a steep learning curve.

To my colleagues, the BiMsters – I consider myself lucky to have been surrounded by a good bunch of people who are intelligent, passionate, inspiring, and most of all, incredibly kind. Tackling one of the world's biggest health problems is no easy feat, but I am proud to have been part of an incredible team that's dedicated to improving the lives of people living with persistent pain. Over the years, we've shared many laughs, words of encouragement and 'Thirstdays' together – thank you for it all. A special thank you to Hayley Leake and Brian Pulling who have been there from the beginning; from my very first day in the lab, you have been the ultimate cheer-squad and have always been willing to lend a listening ear from places both near and far.

Thank you to the participants who volunteered their time to take part in this research – your contributions have filled important gaps in scientific knowledge. Multiple reviewers generously shared their time and wealth of knowledge to improve the quality of this work – I am grateful for their constructive feedback, contributions, and collaboration.

Finally, I would like to acknowledge the support and sacrifices of my loved ones – my family and friends. Thank you for celebrating my triumphs, offering comfort in my times of struggle, and for being patient when I needed to take a social hiatus. I am deeply indebted to each of you. To my parents, Kim and Richard – thank you for your endless love and support, which have been the driving forces behind all of my achievements in life. You have taught me that great things don't come from comfort zones. I love you both so much and I cannot thank you enough for every opportunity you have both afforded me. To my grandparents, Kaye (and Wes), Margaret and John – thank you for the countless times you beamed with pride, or expressed how proud you are of me. To my best friend, Mel – thank you for keeping me grounded and for ensuring that I stepped

outside of the academic sphere every once in a while; I couldn't ask for a more supportive, patient or loving friend. To my second family, the Gavini's – thank you for welcoming me into your home with open arms and for your never-ending love and support. To my nephew, Vinnie – you have a been a ray of light during the final stages of my PhD and you have brought so much happiness into my life. As you grow, I look forward to teaching you about the science of pain. And a heartfelt thank you to my partner, Mitchell. You have played a pivotal role in allowing me to pursue my dreams. Your support, encouragement, belief in me, and undying love have undoubtedly contributed to me being able to reach this achievement. My love for you has no limit, and I would not have wanted to do this without you. I am so looking forward to accomplishing our goals, spending more time together, and seeing how this next chapter of our lives unfolds.

Acknowledgment of funding support:

Candidature support was provided by the University of South Australia on behalf of the Australian Federal Government (Department of Education, Skills and Employment) through an Australian Government Research Training Program Scholarship.

List of Figures

CHAPTER ONE

Figure 1: Engel's biopsychosocial model of health……4

Figure 2: Loeser's onion skin model of pain……5

Figure 3: A depiction of the WCST……...20

CHAPTER TWO

Figure 1: An example of the WCST display

Figure 2: Examples of various scoring scenarios for the WCST according to the Heaton et al. (1993) scoring manual

Figure 3: A Venn diagram illustrating the relationship between perseverative responses and errors made on the WCST

CHAPTER THREE

Figure 1: PRISMA flow diagram

Figure 2: Number of included studies that used each self-report test of cognitive flexibility

Figure 3: Number of included studies that used each neuropsychological test of cognitive flexibility

Figure 4: Correlations of all test pairs accounting for dependency of effect sizes (multi-level random-effects meta-analysis)

Figure 5: Meta-analysis of pooled cohorts that used the shift subscale of the BRIEF-A and the TMT – part B (time)

Figure 6: Meta-analysis of pooled cohorts that used the CFS and the TMT – part B (time)

CHAPTER FOUR

Figure 1: PRISMA flow diagram

Figure 2: The number of included studies that used each self-report test of cognitive flexibility

Figure 3: The number of included studies that used each neuropsychological test of cognitive flexibility

Figure 4: Multilevel random-effects meta-analysis: Correlations of all test pairs accounting for dependency of effect sizes

CHAPTER SIX

Figure 1: An illustrative example of the WCST……..41

Figure 2: Participant flow diagram……..49

CHAPTER SEVEN

Figure 1: Steve Berry's original diagrammatic procedure for scoring unambiguous perseverative responses on the Wisconsin Card Sorting Test…..……....93

Figure 2: Steve Berry's original diagrammatic procedure for scoring unambiguous perseverative responses on the Wisconsin Card Sorting Test (continuation of Figure 1)……..94

Figure 3: Steve Berry's original diagrammatic procedure for scoring ambiguous perseverative responses on the Wisconsin Card Sorting Test…….95

Figure 4: Adaptation of Steve Berry's 1996 diagrammatic procedure for scoring perseverative responses (and perseverative errors) on the Wisconsin Card Sorting Test……96

List of Tables

CHAPTER TWO

Table 1: Key WCST scoring terminology derived from the revised WCST manual (Heaton et al., 1993)

Table 2: A checklist of recommendations for using and reporting the WCST

CHAPTER THREE

Table 1: Study characteristics of included studies

Table 2: Assessment details and correlational data of included studies

Table 3: Reported definitions of cognitive flexibility for included studies

Table 4: Risk of bias assessment

CHAPTER FOUR

Table 1: Risk of bias assessment

Table 2: Study characteristics of included studies

Table 3: Assessment details and correlational data of included studies

CHAPTER SIX

Table 1: Participant characteristics……..50

Table 2: Final regression model for the CFI (proportion of the total score)…….53

Table 3: Final regression model for the WCST (proportion of % perseverative responses)..53

Table 4: Spearman's rho correlation coefficients between self-reported cognitive flexibility and neuropsychological cognitive flexibility…….54

Table 5: Model parameter estimates of each logistic regression model fit in the backward selection process……..55

Preface

This thesis is submitted as a Thesis Containing Published Research in accordance with the University of South Australia Academic Regulations for Higher Degrees by Research Clause 16.1.1.b. It contains a total of nine chapters and includes a combination of both published manuscripts and conventional thesis chapters. Chapters 2-5 consist of inserted manuscripts that have been published in peer reviewed scientific journals. Chapter 6 has been submitted to a journal for peer review and has been formatted to meet the journal requirements. Each embedded published manuscript has independent numbering of pages, figures and tables, and a separate reference list. Where a published manuscript is embedded, the thesis page sequence is momentarily ceased, but resumes following each manuscript. Where the statistical analysis plan and any deviations (supplementary material for Chapter 6; Appendix D) are embedded, the thesis page sequence also ceases briefly, but resumes following its insertion.

Overall, this thesis set out to understand how cognitive flexibility might play a role in persistent pain. **Chapter 1** introduces this thesis. It provides a broad overview of the relevant literature on persistent pain and cognitive flexibility, and outlines both the overarching and specific aims of this thesis. **Chapter 2**, published in *Behavior Research Methods* (impact factor: 5.4; 58 citations at the time of submission), is a critical review that covers the issues with the current use of a popular neuropsychological assessment of cognitive flexibility – the Wisconsin Card Sorting Test (WCST) – particularly with respect to inconsistent terminology and scoring methods. **Chapter 3**, published in *Clinical Psychology Review* (impact factor: 12.8; 23 citations at the time of submission), is a systematic review and meta-analysis that evaluated the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. **Chapter 4**, published in *Neuropsychology* (impact factor: 2.4; 7 citations at the time of submission), is a systematic review and meta-analysis that investigated the relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts. **Chapter 5**, published in the *Australian Journal of Psychology* (impact factor: 2.316; 1 citation at the time of submission), is a critical review that discusses how self-report and neuropsychological assessments of cognitive flexibility have become conflated over time. **Chapter 6** is an online, cross-sectional, observational study that investigates self-report and neuropsychological outcomes of cognitive flexibility in people with and without persistent pain, which was under review at *The British Journal of Pain* at the time of submission. **Chapter 7** is the combined discussion, which offers a summary of the main findings, implications for research and clinical practice, strengths and limitations, unforeseen challenges,

and recommendations for future research. **Chapter 8** contains full details of publications, conference presentations, and awards received throughout the candidature. **Chapter 9** contains all relevant appendices, which include the supplementary material that accompanied all manuscripts, permission statements granted from publishers, and statements of authorship.

“All our dreams can come true, if we have the courage to pursue them”
– Walt Disney

CHAPTER ONE

Introduction

1.1 Pain

1.1.1 *Definition*

Pain is a complex phenomenon (Gatchel, McGeary, McGeary, & Lippe, 2014) that is typically classified as either i) acute or ii) chronic (herein referred to as *persistent* pain). In 2020, the International Association for the Study of Pain (IASP) revised the universally accepted definition of pain for the first time since 1979, with the new definition describing pain as *"an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage"* (Raja et al., 2020). The former IASP definition of pain – "*an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage*" (IASP, 1979) – has endured much scrutiny (Anand & Craig, 1996; Williams & Craig, 2016), but counterarguments have defended it and contended that previous critics have simply misinterpreted the 1979 taxonomic definition (Aydede, 2019). In light of these discussions, the IASP's definition underwent significant review and modifications to address the shortcomings that had been noted. The revised definition is accompanied by extensive explanatory notes that better captures i) pain as a subjective and personal experience, ii) non-human animals and people who are unable to verbalise their pain because of impaired cognitive or language abilities, and iii) modern conceptualisations of pain, thus providing a more explicit and precise understanding of pain for clinicians, researchers, policymakers and persons living with pain.

1.1.2 *Pain and nociception*

Pain is often conflated with nociception, but pain and nociception are in fact distinct phenomena (Wall & McMahon, 1986). While pain is a conscious experience or feeling, nociception is activity in a certain class of neurones (referred to as *nociceptors*) that, under normal conditions, respond to thermal, mechanical, and chemical stimuli (Gatchel, Peng, Peters, Fuchs, & Turk, 2007). Information relating to an actual or potential threat to the tissues is transformed by these neurones and transmitted to the spinal cord. Here, processing of a range of incoming signals may result in a subsequent nociceptive signal to the brain. It is at this point that the experience of pain may or may not be produced (Gatchel et al., 2007). What determines whether or not pain is experienced is highly complex, with some contemporary ideas encompassing every possible cue related to danger or safety, or in other words, the level of perceived threat (Moseley & Butler, 2017). That is, pain emerges once the brain has gathered credible evidence that the body is in danger and requires protection (Butler & Moseley, 2013). This emphasis on the purpose of pain as

a protective mechanism has underpinned contemporary consumer-focused definitions of pain such as 'an unpleasant feeling somewhere in the body that compels a person to protect that body part' (Butler & Moseley, 2013). Although pain and nociception are two processes that can indeed co-exist, it is important to recognise that pain can be experienced in the absence of tissue damage and alternatively, tissue damage can occur without experiencing any pain. Two famous case studies have provided evidence that the relationship between pain and tissue damage is unpredictable (Dimsdale & Dantzer, 2007; Fisher, Hassan, & O'Connor, 1995), with *context* and *meaning* being key moderators. In 1995, the British Medical Journal published the case of a construction worker who presented to the emergency department with excruciating pain after landing on a 15 cm nail that pierced through his boot; however, the nail passed between his toes and never punctured the skin (Fisher et al., 1995). That is, pain was felt in the absence of tissue injury. In 2007, Psychosomatic Medicine published a description of the opposite scenario, that is, no pain in the presence of tissue injury: a construction worker was unaware that he had accidentally shot himself in the face with a nail gun; he went to the dentist six days later with a toothache, and an X-ray revealed that a nail was situated within his skull. Experimental studies have corroborated these findings by demonstrating the importance of contextual information in the experience of pain. A very cold noxious stimulus evokes more intense pain when accompanied with implicit (Moseley & Arntz, 2007) or explicit (Arntz & Claassens, 2004) cues – despite the temperature of the stimulus remaining unchanged – because of the meaning attributed to indicators of perceived danger relative to safety.

1.1.3 *When does acute pain transition into persistent pain?*

Generally, acute pain associated with physical injury tends to subside once the injured bodily tissue has had sufficient time to heal and pain is no longer required to reduce exposure of the body part to mechanical load. However, when pain continues or recurs beyond the typical timeframe required for healing after sustaining an injury i.e., 3 months or longer (Merskey & Bogduk, 1994; Treede et al., 2015), pain is characterised as persistent because it can no longer be explained by tissue damage or inflammation.

1.1.4 *The biopsychosocial model of pain*

The biopsychosocial model, which was first conceptualised by George Engel in 1977 (Engel, 1977), posited that in order to understand various health conditions, consideration must be given to the complex interaction between biological, psychological, and social factors that form an integral part of an individual's clinical presentation (see Fig 1). Until the end of the 20th century,

understandings about disease and illness remained deeply seated within the *biomedical* model – an approach that relies on pathoanatomical explanations for disease and illness, which was, at the time, considered to be the pinnacle of 'objectivity'. However, the biomedical framework has failed to adequately explain conditions where no pathoanatomical cause can be identified, which is the case for many instances of persistent musculoskeletal pain (Lewis & O'Sullivan, 2018). Rather, pain and tissue pathology share a variable relationship. For example, previous studies reveal that approximately 35% to 40% of elite athletes have a bulged disc on magnetic resonance imaging (MRI) without a history of back pain (Alyas, Turner, & Connell, 2007; Bezuglov et al., 2021) and associations between lumbar spine MRI findings and outcomes for low back pain remain inconsistent (Steffens et al., 2014). Comprehensive systematic reviews of randomised controlled trials (RCTs) have consistently indicated that invasive procedures, which align with the biomedical model, are no more effective than sham treatments for persistent pain conditions (Jonas et al., 2019; Louw, Diener, Fernandez-de-las-Penas, & Puentedura, 2017). The reductionist and exclusionist approach of the biomedical model in understanding and treating health conditions limited the contribution of psychosocial factors in health and illness, and in turn, these components have historically received less attention in the medical sphere. However, Engel's seminal paper (Engel, 1977) arguably shifted scientific enquiries and clinical practice towards a broader approach to healthcare that not only integrates the biological components, but also psychological and social factors into an all-encompassing person-centered model.

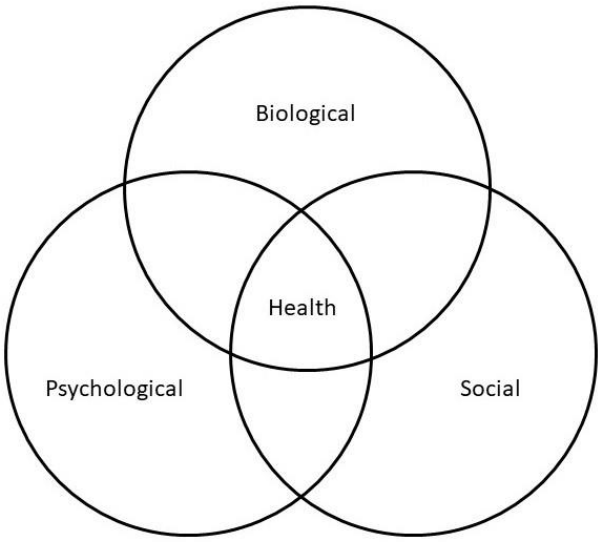


Figure 1. Engel's biopsychosocial model of health (Engel, 1977). Contemporary understandings of pain use '*pain'* in lieu of '*health'*.

Building on Engel's biopsychosocial model (Engel, 1977), John Loeser's conceptual model of pain (Loeser, 1982), otherwise known as the 'onion skin model' of pain and suffering (see Fig 2), is renowned for recognising that sensory, affective, and social factors can influence, and be influenced by, pain and suffering. Specifically, the model suggests that there are four components of pain itself: nociception, pain, suffering, and pain behaviours (Loeser, 2000). Although aspects of Loeser's model are no longer consistent with contemporary understandings of pain – particularly his idea that pain only occurs without nociception when the nervous system is damaged (Loeser, 2000) – the model has nonetheless vastly contributed to present understandings and treatment of pain (Bevers, Watts, Kishino, & Gatchel, 2016).

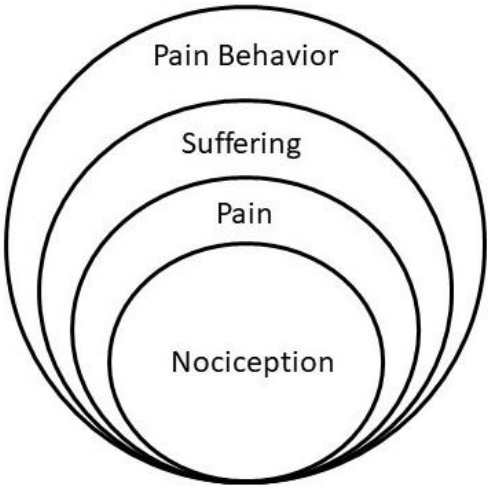


Figure 2. Loeser's onion skin model of pain (Loeser, 1982). Reproduced with permission from Wolters Kluwer.

The biopsychosocial model has made important contributions to the pain field as it fundamentally changed how persistent pain is understood and treated. It provides a comprehensive framework for understanding how biological, psychological, and social factors collectively influence the experience of pain – either acute or persistent (Gatchel et al., 2007) – and is geared toward reframing pain as 'the *perceived* need to protect the body from harm' rather than from a biomedical perspective that has historically conceptualised pain as 'an accurate marker of the extent of tissue damage' (Moseley, 2003; Moseley & Butler, 2015). The biopsychosocial model is currently deemed to be the most appropriate and efficacious approach for the evaluation and treatment of persistent pain (Gatchel et al., 2007; Louw, Schmidt, Zimney, & Puentedura, 2019;

Meeus et al., 2016; Moseley, 2002), with recommendations for best practice emphasising the importance of providing care that is individually tailored, acknowledges that the mechanisms underpinning persistent pain are multifaceted, and directly involves the patient in their recovery (Lin et al., 2020). Despite this, however, many barriers have the potential to prevent the uptake and implementation of such evidence-based interdisciplinary care (Gatchel et al., 2014; Meeus et al., 2016; Moseley, 2003), and a large proportion of patients (43%) seeking primary care for conditions such as low back pain report low satisfaction levels with their treatment (Ahern, Dean, Dear, Willcock, & Hush, 2019).

1.1.5 *Prevalence, burden and risk factors*

Persistent pain is common. Large-scale epidemiological studies and systematic reviews have estimated that the point prevalence of persistent pain ranges between 12% to 51.3% (Blyth et al., 2001; Fayaz, Croft, Langford, Donaldson, & Jones, 2016; Nahin, 2015). The considerable lack of precision in incidence estimates may be attributed to the inconsistencies that are often observed across studies, particularly with regards to definitions of persistent pain, the use of different methodological approaches, or sociodemographic factors of the studied population. Despite these disparities, the general consensus is that persistent pain affects approximately 20% of the general adult population in both developed (Blyth et al., 2001; Breivik, Collett, Ventafridda, Cohen, & Gallacher, 2006; Pitcher, Von Korff, Bushnell, & Porter, 2019) and developing countries (Sá et al., 2019), making it a major public health concern worldwide. Furthermore, persistent pain has an estimated annual incidence rate of 8.3% and a recovery rate of 5.4% (Elliott, Smith, Hannaford, Smith, & Chambers, 2002), with the lifetime prevalence of any episode of low back pain, for example, being as high as 80% (Walker, 1999; Walker, Muller, & Grant, 2003). Taken together, the evidence from epidemiological research demonstrates the extent of the impact of pain and persistent pain conditions.

Persistent pain exerts a substantial burden from both an individual and societal standpoint. In 2018, it was estimated that 3.24 million Australians had persistent pain, with that figure expected to escalate to 5 million people by 2050 – a value that is consistent with the ageing population and increased life expectancy (Deloitte Access Economics, 2019). Several studies have indicated that pain is a common reason for utilising health care services (Blyth, March, Brnabic, & Cousins, 2004; Igumbor, Puoane, Gansky, & Plesh, 2012; Mäntyselkä et al., 2001; St. Sauver et al., 2013), most likely because pain is a defining feature of many chronic conditions. In 2018, the annual expenditure for persistent pain was estimated at over $73 billion in Australia ($139 billion if

impacts on quality of life were considered) (Deloitte Access Economics, 2019), and in 2010, direct and indirect costs totalled between $560 and $635 billion in the United States – exceeding the combined cost of cardiovascular disease and cancer (Gaskin & Richard, 2012).

According to the most recent Global Burden of Disease Study, persistent pain conditions have remained in the top four leading causes of years lived with disability worldwide since 1990 (James et al., 2018) and consequently, individuals with persistent pain often experience detrimental effects on various aspects of their physical, psychological and social functioning (Turk, Fillingim, Ohrbach, & Patel, 2016). The combination of modifiable lifestyle factors such as physical activity, smoking status, alcohol consumption, and opioid use have collectively been shown to account for at least 53.4% of the association between persistent pain and all-cause mortality (Chen et al., 2021). Persistent pain also has a vast impact on workforce absenteeism and reduced work productivity. One study estimated that the combined annual cost to Australian employers for pain-related presenteeism and absenteeism is 5.1 billion dollars (van Leeuwen, Blyth, March, Nicholas, & Cousins, 2006); however, other studies have reported that presenteeism, but not absenteeism, is associated with the presence of persistent pain (Kawai, Kawai, Wollan, & Yawn, 2017). Perhaps unsurprisingly, pain is a common reason for early departure from the workforce. A large retrospective cross-sectional analysis study that used data from the Australian Bureau of Statistics 2003 Survey of Disability, Ageing and Carers found that 41.4% of Australians aged between 45-64 years were no longer working because of back-related problems (Schofield, Shrestha, Passey, Earnest, & Fletcher, 2008). Another Australian study found that the total wealth accumulated for people who leave the work force because of back problems is 87% less than for those who are employed full-time with no chronic health condition, after controlling for demographic factors including age, sex, and education (Schofield et al., 2011).

The negative impact of pain extends further when considering that those with persistent pain experience greater interference with daily activities and more sleep disturbances (Kawai et al., 2017; Mäntyselkä et al., 2001), and are less physically active than their healthy counterparts (McLoughlin, Colbert, Stegner, & Cook, 2011). A recent mixed-methods study examining the impact of persistent pain on patients' quality of life found that persistent pain negatively impacted patients' lives in terms of their physical activity, occupation, relationships, social activities, sleep, and mood (Hadi, McHugh, & Closs, 2019), solidifying the results of previous research. Persistent pain is also associated with higher rates of depression and anxiety than that found in the general population (McWilliams, Cox, & Enns, 2003), with comorbid persistent pain and major depressive disorder reducing people's quality of life to a larger extent than either condition alone (Arnow et

al., 2006). Pain and mental health conditions such as depression and anxiety are not only highly comorbid, but they probably share a causal bidirectional relationship; the incidence of one condition increases the risk of the other during a 10-year period, but the exact mechanisms behind the onset of each condition remain to be elucidated (Bondesson et al., 2018).

There are many modifiable and non-modifiable risk factors that have the potential to predispose individuals to develop persistent pain (Mills, Nicolson, & Smith, 2019; van Hecke, Torrance, & Smith, 2013). Persistent pain is a condition that affects all age groups across the lifespan from adolescence to older adulthood, but tends to be more common in older age (i.e., between 40-80 years; Bergman et al., 2001; Breivik et al., 2006; Hoy et al., 2012). Many studies have consistently found that persistent pain is more prevalent among women than men (Bergman et al., 2001; Breivik et al., 2006; Fayaz et al., 2016; Hoy et al., 2012; Mogil, 2012; Nahin, 2015). Racial and/or ethnic background is an important non-modifiable risk factor for the development of persistent pain, with minority groups being more likely to experience persistent pain, report greater pain intensity and disability, and experience poorer outcomes concerning treatment than racial and ethnic majorities (Campbell & Edwards, 2012). A recent study examining the effect of race on persistent pain found that White adults experience less pain intensity than their Black and Hispanic counterparts, and also experience less pain-related disability in certain aspects of their lives than Black and Hispanic individuals (Janevic, McLaughlin, Heapy, Thacker, & Piette, 2017). A UK study that surveyed over 500,000 participants found that persistent pain was associated with socioeconomic factors such as lower education, lower annual income, and being unemployed, as well as psychosocial factors such as the number of depressive episodes and adverse life events (Macfarlane, Beasley, Smith, Jones, & Macfarlane, 2015). Further, this study also found that those who self-identified as White were less likely to report persistent pain than those who self-identified as Black, Asian or mixed ethnicity; these group differences remained, albeit attenuated, after adjusting for income, employment status and number of adverse life events (Macfarlane et al., 2015). Together, these studies highlight the influential role of socioeconomic and psychosocial factors in persistent pain conditions. In acute settings, post-operative pain severity has also been found to differ among different ethnic groups in patients undergoing a surgical procedure (Faucett, Gordon, & Levine, 1994).

Modifiable risk factors associated with persistent pain include, but are not limited to, lifestyle factors such as smoking, excess alcohol consumption, extreme Body Mass Index values (<18.5 and >30), sleep disturbances, and mental health factors (Mills et al., 2019; van Hecke et al., 2013). Psychosocial factors such as depression, fear-avoidance beliefs, and pain catastrophising are

predictive of chronicity and pain-related disability (Hruschak & Cochran, 2018; Turk et al., 2016). However, it is important to consider why certain individuals are more susceptible to the development, maintenance, and recovery of persistent pain than others – identifying specific mechanisms underlying these processes remains one of the biggest mysteries in the field of pain.

1.2 When pain persists – chronic pain and learning

Persistent pain is associated with a range of processes that involve neural, immune, endocrine, and autonomic systems. Differences have been demonstrated between those with and without persistent pain in stimulus-response profiles from across the sensory neural axis (Latremoliere & Woolf, 2009). Abnormal brain function, structure, and chemistry have also been demonstrated in people with persistent pain (Martucci & Mackey, 2018; Wand et al., 2011). Each of the changes that occur within the nervous system can be broadly categorised as learning. That is, a large body of scientific evidence suggests that pain can be (un)learned because of the innate malleability of the nervous system, providing a potential explanation for pain chronification and recovery (Butler & Moseley, 2013; Moseley & Butler, 2017). While several theoretical models have been put forward to offer potential explanations for the development and maintenance of persistent pain, three prominent theories are particularly relevant to this thesis. The most studied and most established form of learning within the context of persistent pain is long-term potentiation. More recently, frameworks such as classical conditioning and predictive processing have also been applied to understand the potential mechanisms that may underlie the transition from acute to persistent pain, its perseveration and the capacity to recover, and how people become more sensitive to noxious and non-noxious stimuli as pain persists.

1.2.1 *Long-term potentiation – 'neurones that fire together, wire together'*

The discovery of long-term potentiation (LTP) – a form of learning that refers to the enhanced efficacy of a neural synapse (McCarberg & Peppin, 2019) – was first clearly demonstrated (Bliss & Collingridge, 1993; Bliss & Lømo, 1973) 24 years after the phrase 'neurones that fire together, wire together' (Hebb, 1949). The most common trigger for LTP is the increased signalling frequency of the pre-synaptic neurone, which triggers a sophisticated cascade of events within the post-synaptic neurone. These events include the opening of ion channels by neurotransmitters, the opening of separate voltage-driven ion channels, and the production and insertion of extra ion channels on the post-synaptic membrane. All these events together enhance the responsiveness of the post-synaptic neurone and thus enhance synaptic efficacy. It is the latter stages of these events – the production of extra ion channels and their insertion in the cell wall, that are thought

to mediate LTP. With respect to persistent pain, this form of learning occurs throughout the network of neurones that are involved in processing nociceptive inputs (Milligan & Watkins, 2009). One effect of LTP is that the stimulus intensity required to activate nociceptive networks within the spinal cord and brain is reduced, which often manifests as allodynia, or a reduced pain threshold – that is, pain is experienced on previously non-painful events. Another effect of LTP is that stimuli that previously activated nociceptive networks now have a larger activating effect, which often manifests as hyperalgesia, or 'more pain for normally painful events' (Moseley & Butler, 2017).

1.2.2 *Associative learning*

Associative learning, the most well-known of which is classical or Pavlovian conditioning (Pavlov, 2010), involves the development of an association between two stimuli such that both can evoke the response previously only evoked by one. That is, when a biologically salient unconditioned stimulus (US) becomes associated with a formerly neutral stimulus (NS), it becomes a conditioned stimulus (CS). A CS elicits the response formerly elicited by the US, and the response is now termed a 'conditioned response' (CR) (Clark, 2004). Anecdotally, associative learning *seems* a likely mechanism for the development of pain in situations that are not dangerous to body tissue (Madden & Moseley, 2016). However, two systematic reviews demonstrated that there was insufficient experimental evidence to show associative learning effects on pain (Harvie, Moseley, Hillier, & Meulders, 2017; Madden et al., 2015). Since then, several experiments have demonstrated classically conditioned allodynia (Harvie et al., 2015; Madden et al., 2016; Traxler, Madden, Moseley, & Vlaeyen, 2019) and classically conditioned hyperalgesia (Harvie, Meulders, Madden, et al., 2016), but classically conditioned pain in the absence of nociceptive stimulation, remains to be clearly demonstrated. The role of associative learning in the context of chronic pain has received renewed interest since the publication of the Imprecision Hypothesis (Moseley & Vlaeyen, 2015). That hypothesis postulated that conditioned pain, allodynia, and hyperalgesia generalises to novel stimuli that have never been dangerous, and that the extent of generalisation is inversely dependent on the precision with which the non-nociceptive characteristics of the original compound (or complex) stimulus, is encoded by the brain. Although there are some exceptions (Madden, Russek, Harvie, Vlaeyen, & Moseley, 2017), experiments have supported aspects of the hypothesis (Chaves et al., 2021; Harvie, Meulders, Madden, et al., 2016; Harvie, Meulders, Reid, et al., 2016; Meulders et al., 2014), but it remains to be fully interrogated.

1.2.3 *Predictive processing: What we expect influences what we get*

Conventional models of perception have typically viewed the brain as a 'stimulus-response' organ, whereby neuronal activity passively awaits the arrival of a modulatory stimulus. In contrast to this view, predictive processing provides a contemporary and compelling conceptual framework for understanding the inner workings of the human brain (see Clark, 2016 for a comprehensive overview of sensation, perception, and action), particularly in terms of its structure and function. Further, it provides a sensible explanation for how the brain plays an active role in processing and combining sensory and contextual information to construct a coherent percept (Friston, 2005, 2009; Hohwy, 2013, 2020; Rao & Ballard, 1999). The theory posits that the brain, or 'agent' itself, is an organism that tries to anticipate, in an ongoing cyclic fashion, the everchanging regularities of the internal (i.e., interoception) and external (i.e., exteroception) world. This is said to be achieved by the brain formulating predictions based on previous beliefs, knowledge, expectations, and experience (otherwise termed 'priors') that are combined with incoming sensory inputs (Clark, 2016; Hohwy, 2020).

Predictive processing has recently become a highly influential theory in the area of pain research and has been put forward as a potential explanation underpinning the chronification of pain (Ploner, Sorg, & Gross, 2017; Tabor & Burr, 2019; Tabor, Thacker, Moseley, & Kording, 2017; Wiech, 2016) and placebo effects (Büchel, Geuter, Sprenger, & Eippert, 2014; Kaptchuk, Hemond, & Miller, 2020; Ongaro & Kaptchuk, 2019). Any discrepancy that is detected between the predicted and the actual sensory signal will produce a *prediction error*, which is then transmitted along forward connections towards higher hierarchical levels to refine predictions, optimise the accuracy of the internal model, and minimise prediction error in the future (Clark, 2016; Friston, 2010; Hohwy, 2013). Prediction errors can be resolved in one of two ways i) by adjusting the predictions that are generated by the generative model to better match the sensory input, or ii) by holding the predictions of the generative model constant and altering the sensory input via action (two contrasting perspectives on the topic of action exist: active inference as discussed by Clark, 2016; or the action prevents error thesis as proposed by Vance, 2017).

Occasionally, however, hierarchical generative models have the potential to mislead (Clark, 2016) or 'failures of inference' distort the perceptual experience (Henningsen et al., 2018), for example: What I if told you that you read this sentence wrong? This depicts how the brain actively tries to predict ambiguous upcoming sensory input based on prior knowledge (*read that example again*). Yet, *priors* are capable of overriding parts of the sensory signal to create a percept that is

'optimal' to minimise prediction error (Clark, 2016). In the context of persistent pain, it seems reasonable to speculate that if the brain and nervous system hold a strong prior that pain will occur upon bending forward, for example, then non-nociceptive signals (i.e., benign movements) will be interpreted as 'noise', and the perceptual experience will be shifted in favour of the prediction of persisting pain, despite no clear indication of tissue pathology (Kaptchuk et al., 2020). In other words, the more stable the generative model, the less influence prediction error has on updating the model. It is argued that the brain relies on its innate computational architecture to approximate Bayes theorem (see Knill & Pouget, 2004; Parr, Rees, & Friston, 2018) using probabilistic density functions. That is, perception is dependent on the relative weighting or 'precision' of predictions and prediction errors, respectively (Friston, 2010).

Probabilistic inference requires the reciprocal exchange of top-down and bottom-up information in a flexible manner, which is thought to play an important role in facilitating learning (Clark, 2016). A few experimental studies have demonstrated that when noxious or non-noxious stimuli are associated with contextual cues that represent *danger* or *safety,* pain can be evoked in the absence of clear tissue pathology in people with (Harvie et al., 2015) and without persistent pain (Bayer, Baer, & Early, 1991; Moseley & Arntz, 2007). These findings raise the possibility that non-resolution of pain when nociceptive signalling has resolved may reflect disruption of this error detection and model updating system. That is, perhaps persistence of pain after an acute episode reflects a failure of the system to respond to changing sensory inputs – a form of inflexibility in a changing sensory environment.

1.3 Cognitive flexibility

1.3.1 *Definition*

Cognitive flexibility, or terms that are often conceptually thought to reflect the opposite of cognitive flexibility (i.e., *in*flexibility or rigidity; Diamond, 2013), have a long history grounded in the field of (neuro)psychology that dates back to the mid-to-late 20th century (Schultz & Searleman, 2002; Scott, 1962; Stewin, 1983). Defining the construct of cognitive flexibility has proven to be difficult, at least in part, because of the various thoughts and/or behaviours that are considered to be 'flexible', the plethora of assessments that are used to assess cognitive flexibility, and the lack of definitional consensus among experts in the field (Cherry, Vander Hoeven, Patterson, & Lumley, 2021; Ionescu, 2012). The construct of cognitive flexibility was described as a source of confusion to researchers and clinicians almost four decades ago (Stewin, 1983). Although there is no unanimous definition, there is general agreement that cognitive flexibility can

be broadly described as the ability to adapt to changing task demands or environmental conditions by selectively adjusting one's thoughts, cognitive strategies, and/or behavioural responses (Cañas, Quesada, Antolí, & Fajardo, 2003; Deák, 2003; Scott, 1962). Some ways in which cognitive flexibility can manifest include changing perspectives, generating different ideas, altering responses when conditions change, and/or shifting thoughts and behaviours to respond in ways that are adaptive (Diamond, 2013; Eslinger & Gratten, 1993).

Executive function is the umbrella term used to define higher-order abilities that facilitate the regulation of thoughts and actions during goal-directed behaviour (Friedman & Miyake, 2017). Some researchers have proposed that there are three separate components of executive function: cognitive flexibility, working memory and inhibition (Diamond, 2013) or set-shifting, updating and inhibition (Miyake, Friedman, et al., 2000), while others have indicated that cognitive flexibility is an emergent property of executive function (Ionescu, 2012) thought to derive from the complex coordination of high-level cognitive processes such as working memory, inhibition, switching and attention (Dajani & Uddin, 2015; Deák, 2003). More recently, the 'variability-stability-flexibility' pattern of cognitive flexibility has also been proposed (Ionescu, 2017). Thus, whether cognitive flexibility is a distinct component, or an emergent property of executive function, is a point of ongoing debate.

Often cognitive flexibility is operationalised by set-shifting or task switching tasks (Dajani & Uddin, 2015). Although the terms set-shifting and cognitive flexibility have been used synonymously (Ionescu, 2012), set-shifting – a lower-level component of cognitive flexibility – refers to the ability to shift between different rule sets to complete a task (Dajani & Uddin, 2015; Tchanturia et al., 2004). On the other hand, task-switching is thought to be a higher-level and more complex form of cognitive flexibility that involves regularly switching between two or more tasks, each with their own set of instructions (Badre & Wagner, 2006; Dajani & Uddin, 2015). For instance, during a set-shifting task the rules about the specific characteristics of the stimuli change within the task (i.e., sorting a stimulus according to colour, then form), but the instructions of the task (i.e., correctly sort the stimulus) are held constant, whereas task-switching paradigms may involve, for example, determining whether a number is odd or even, or if a letter is a vowel or consonant for letter-number combinations (e.g., 3c). Task-switching tasks are more complex than set-shifting tasks because the instruction cue about how to accurately categorise stimuli are repeated for one half of the trials (e.g., "LETTER" then "LETTER"), while the other half involves a task-switch (i.e., "LETTER" then "NUMBER") (Badre & Wagner, 2006). Cognitive and behavioural manifestations of what is considered to be cognitively inflexible can be both advantageous and detrimental under certain

circumstances (Uddin, 2021) and cognitive flexibility, as indexed by task-switching paradigms, can also be positively reinforced (via rewards) and be influenced by implicit contextual cues (Braem, 2017; Braem & Egner, 2018).

1.3.2 *This thesis is not about psychological flexibility*

It is important to differentiate *psychological flexibility* from *cognitive flexibility*. Although psychological and cognitive flexibility are two constructs that originated from different theoretical backgrounds, the terms have been used interchangeably in the literature (e.g., Asali et al., 2020; Palm & Follette, 2011) despite the fact that only one study to date has empirically examined the relationship between the two constructs (Grant & Cassidy, 2022).

Psychological flexibility, defined as "the ability to contact the present moment more fully as a conscious human being, and to change or persist in behavior when doing so serves valued ends" (Hayes, Luoma, Bond, Masuda, & Lillis, 2006, p. 7) is central to Acceptance and Commitment Therapy (ACT; Hayes, Strosahl, & Wilson, 1999) – a key therapy in the so-called 'third wave' cognitive and behavioural therapies (CBT). Unlike traditional CBT models, which place a large amount of emphasis on having to change thoughts and cognitions directly to result in behaviour change and clinical improvement (Hayes et al., 2006), ACT is described as having theoretical grounds in behavioural psychology and mindfulness and acceptance-based approaches, and was developed to be a contextualised and more practical therapeutic intervention that aims to change the relationship between cognitions and behaviours (i.e., avoiding interpreting mental events as verbatim). According to this model, six-core processes are thought to foster *psychological flexibility*: acceptance, cognitive defusion, contact with the present moment, self as context, values, and committed action (Hayes et al., 2006).

Over the last two decades, psychological flexibility, referred to as a "fresh new concept" (Vlaeyen, 2014, p. 235), is a construct that has been extensively investigated, particularly in people with persistent pain. Psychological flexibility, as assessed by the Psychological Inflexibility in Pain Scale (PIPS) and the Valuing Questionnaire (VQ), has been shown to be a resilience factor among people with persistent pain (Gentili et al., 2019) and changes in the core processes underlying psychological flexibility often predict optimal outcomes with physical, psychological, and social functioning (Chisari et al., 2022; McCracken & Gutiérrez-Martínez, 2011; Zetterqvist, Holmström, Maathz, & Wicksell, 2017). Additionally, a recent study found that women who consider themselves to have high psychological flexibility (self-reported on the Psychological Flexibility Questionnaire) were more likely to partake in vaginal delivery over vacuum extraction or caesarean delivery

during labour than women who rated themselves as less psychologically flexible (Asali et al., 2020).

Recently, however, concerns about the construct validity of the most widely implemented assessment of psychological flexibility (the Acceptance and Action Questionnaire II (AAQ-II)) have come to the fore (Doorley, Goodman, Kelso, & Kashdan, 2020), with research demonstrating that the construct is highly correlated with emotional distress (Tyndall et al., 2019; Wolgast, 2014). A recent commentary of Chisari (2022) by Chalmers (2022) duly noted that using non-validated tools to assess psychological flexibility in women with vulvodynia, as well as only assessing a subset of the components of the construct (i.e., pain acceptance, present moment awareness and committed action), is not sufficient to recommend ACT as a suitable intervention for those who experience vulvodynia. In light of these limitations, Chalmers (2022) suggested that the AAQ-II or the pain-related version (AAQ-II-P) may be a more appropriate assessment to better capture all six components of psychological flexibility. It should be emphasised, however, that the AAQ-II, as well as the original version (AAQ), have never been validated for use within pain populations and according to Reneman et al. (2014), the AAQ-II-P does not specifically assess experiential avoidance in relation to *persistent* pain because none of the items in the questionnaire address pain duration.

What is even more concerning is that other assessments that are often labelled as questionnaires that specifically target 'psychological flexibility' (or other terms that have been largely replaced by *psychological flexibility* such as 'acceptance' or 'experiential avoidance'; Wolgast, 2014) in pain populations have received little-to-no psychometric validation against other assessments designed to assess the same construct. Despite being designed to assess pain acceptance, the Acceptance of Pain Questionnaire (ACPT), which was originally developed by Geiser (1992) and later renamed as the Chronic Pain Acceptance Questionnaire (CPAQ) (McCracken, Vowles, & Eccleston, 2004), seems to have not been validated against other assessments designed to assess the same construct, thus its convergent validity remains unestablished. Moreover, early validation studies such as those by McCracken et al. (2004) did not correlate the revised version of the CPAQ with the AAQ – two assessments that should theoretically share high construct validity. Questionnaires such as the PIPS (Wicksell, Renöfält, Olsson, Bond, & Melin, 2008) are also limited in their scope when it comes to capturing all six components of psychological flexibility because it only has two subscales: *avoidance of pain* and *cognitive fusion with pain,* and thus the assessment excludes other aspects that are central to the ACT model. Together, these issues indicate the breadth of problems that make assessing the construct of psychological flexibility rather

problematic. Suffice here, however, to iterate that although there seems to be common problems, interchangeable usage and common relevance to persistent pain, psychological and cognitive flexibility are different, and further consideration of the former is beyond the scope of this thesis.

1.3.3 *The lack of definitional consensus surrounding cognitive flexibility*

Definitions of cognitive flexibility are far from scarce. Previous research indicates that cognitive flexibility is a construct that is yet to reach a unified understanding among experts in the field and thus, remains poorly defined in the literature – both conceptually and operationally (Dajani & Uddin, 2015; Ionescu, 2012). Indeed, the field has tended to use the term *cognitive flexibility* synonymously with other descriptors such as *psychological flexibility, mental flexibility, set-shifting* (considered to be one component of cognitive flexibility by some researchers e.g., Lounes, Khan, & Tchanturia, 2011), and *switching* (e.g., Asali et al., 2020; Diamond, 2013; Palm & Follette, 2011; Wecker, Kramer, Hallam, & Delis, 2005). A recent scoping review identified similar problems with the interchangeability of terms and definitions surrounding the construct of *psychological flexibility* (Cherry et al., 2021), highlighting that such conceptual issues are not confined to the construct of cognitive flexibility.

1.3.4 *The development and decline of cognitive flexibility across the lifespan*

Despite the apparent definitional and operational problems with cognitive flexibility, there is consensus regarding its developmental trajectory, which follows an inverted U-shape pattern (Cepeda, Kramer, & Gonzalez de Sather, 2001; Kupis et al., 2021). A combination of theoretical models and empirical data indicate that cognitive flexibility undergoes distinct changes across the lifespan, increasing during childhood, peaking and remaining relatively plateaued during adulthood – particularly during 20-30 years of age – and declining in older age (i.e., after 60 years of age; Ashendorf & Mccaffrey, 2008; Cepeda et al., 2001; Schultz & Searleman, 2002; Wecker, Kramer, Hallam, & Delis, 2005).

Age-related changes in cognitive flexibility are likely to share parallels with neurophysiological changes of the brain– both structural and functional – that are not only distinctive of particular stages of development (Buttelmann & Karbach, 2017), but are often observed during the onset of various conditions that impact typical development (e.g., autism spectrum disorder) or normal ageing (e.g., Alzheimer's disease). Such conditions tend to follow sequential transitions across the lifespan and are characterised by cognitive inflexibility (Uddin, 2021). Indeed, there is evidence that specific brain dynamic patterns are associated with changes in cognitive flexibility across the

lifespan (Kupis et al., 2021), supporting the notion that fluctuations in cognitive flexibility that occur throughout life may have a neurophysiological basis.

1.3.5 *Cognitive flexibility and its relation with cognitive abilities, psychopathology and physical aspects of health*

Cognitive flexibility has been associated with favourable academic abilities such as mathematical and reading performance in children (Colé, Duncan, & Blaye, 2014; Yeniad, Malda, Mesman, van Ijzendoorn, & Pieper, 2013), and health-related quality of life outcomes in older adults (Davis, Marra, Najafzadeh, & Liu-Ambrose, 2010). It has been negatively associated with intolerance of uncertainty and perceived stress among university students (Demirtas & Yildiz, 2019). Reduced cognitive flexibility has been associated with several conditions including mild traumatic brain injury (Pang, Dunkley, Doesburg, da Costa, & Taylor, 2016), autism spectrum disorder (Fujino et al., 2019; Westwood, Stahl, Mandy, & Tchanturia, 2016), persistent pain (Bell et al., 2018; Berryman et al., 2014; Bunk, Preis, Zuidema, Lautenbacher, & Kunz, 2019), eating disorders, particularly anorexia nervosa (Lounes et al., 2011; Steinglass, Walsh, & Stern, 2006; Westwood et al., 2016), schizophrenia (Thoma, Wiebel, & Daum, 2007), comorbid anxiety and depression (Johnco, Wuthrich, & Rapee, 2014) as well as major depressive disorder alone (Lee, Hermens, Porter, & Redoblado-Hodge, 2012).

Taken together, the evidence suggests that cognitive flexibility is a core feature of cognitive development, well-being, and psychopathology, and appears to be an important target for treatment. Cognitive remediation therapy – a popular intervention for the treatment of eating disorders and autism spectrum disorder – aims, in part, to improve cognitive flexibility (Okuda et al., 2017; Tchanturia, Lloyd, & Lang, 2013) and neuropsychological interventions have also been used to enhance cognitive flexibility, with favourable outcomes reported in those with post-traumatic stress disorder symptoms (Ben-Zion et al., 2018).

1.3.6 *Assessing cognitive flexibility*

To date, an abundance of assessments have been developed and used to evaluate the construct of cognitive flexibility in both research and clinical settings. There are two broad approaches to assessing cognitive flexibility: self-report questionnaires and neuropsychological assessments.

1.3.6.1 *Neuropsychological assessments*

Traditionally, so-called *objective* performance-based tests – herein referred to as neuropsychological assessments (or 'tasks' or 'tests') – have been used to assess cognitive flexibility and such modes of neuropsychological testing are currently argued to be the 'gold standard' (Johnco et al., 2014). Neuropsychological assessments such as the Wisconsin Card Sorting Test (WCST; Berg, 1948; Grant & Berg, 1948), the Trail Making Test (TMT; Reitan, 1955, 1958), the Controlled Oral Word Association Test (COWAT; Benton, Hamsher, & Sivan, 1994) and the Stroop Colour Word Test (Stroop; Stroop, 1935) have not only been ranked among the top 40 most frequently used instruments administered by clinical neuropsychologists (Rabin, Barr, & Burton, 2005), but have become widely popularised as neuropsychological assessments of cognitive flexibility (Dennis & Vander Wal, 2010; Johnco et al., 2014; Tchanturia et al., 2012).

Other neuropsychological assessments that are often used to assess cognitive flexibility include the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997), the Intra-Extra Dimensional Task (Robbins et al., 1998), the CatBat task (Eliava, 1964; Tchanturia, Morris, Surguladze, & Treasure, 2002), the Haptic Illusions Task (Tchanturia, Serpell, Troop, & Treasure, 2001; Uznadze, 1966), the Alternate Uses Task (Guilford, 1956) and sub-tests from the Delis-Kaplan Executive Function System (e.g., the Verbal Fluency task) (Delis, Kaplan, & Kramer, 2001).

1.3.6.2 *Self-report questionnaires*

Self-report questionnaires aiming to evaluate the all-encompassing construct of cognitive flexibility were not introduced until the end of the 20th century. Self-report questionnaires typically involve participants indicating their level of agreement or disagreement with statements about cognitive (in)flexibility using a Likert scale. The Cognitive Flexibility Scale (CFS; Martin & Anderson, 1998; Martin & Rubin, 1995) was the first self-report questionnaire to be developed to assess the construct of cognitive flexibility. The 12-item scale was designed to capture three components of cognitive flexibility that are thought to be important for interpersonal communication competence: (a) awareness of alternatives (e.g., "I can communicate an idea in many different ways"), (b) willingness to adapt to situations (e.g., "I have many possible ways of behaving in any given situation") and (c) level of self-efficacy in being flexible (e.g., "I have the self-confidence necessary to try different ways of behaving") (Martin & Rubin, 1995). Self-report assessments are relatively brief and simple to administer and score – making them practical for clinical and research-based settings (Dennis & Vander Wal, 2010; Johnco et al., 2014) – as well as being able to

capture information that is more ecologically representative of real-world scenarios (Isquith, Roth, & Gioia, 2013).

The CFS prompted the development of the Cognitive Flexibility Inventory (CFI; Dennis & Vander Wal, 2010), a questionnaire comprising of 20-items and two subscales – *Alternatives* (13 items; e.g., "I consider multiple options before making a decision") *and Control* (7 items; e.g., "When I encounter difficult situations, I feel like I am losing control"). The CFI was created to assess the components of cognitive flexibility that allows for the restructuring of maladaptive cognitions to be more adaptive during stressful situations, for example, (a) the ability to discern challenging situations as controllable; (b) the ability to perceive alternate reasons for life occurrences and human behaviour; and (c) the ability to produce alternative solutions to difficult situations. In turn, the CFI may afford more utility in clinical practice and be a useful assessment for cognitive-behavioural interventions because of its two-factor structure, and its capacity to be administered to the same individual on different occasions due to high its test-retest reliability (Dennis & Vander Wal, 2010). Previous studies have indicated that self-reported anxiety and depression are negatively associated with self-reported cognitive flexibility (Dennis & Vander Wal, 2010; Johnco et al., 2014; Lounes et al., 2011; Wang et al., 2019; Yu, Yu, & Lin, 2020), but not with performance on neuropsychological assessments of cognitive flexibility (Lounes et al., 2011). One reason for the lack of correlation between outcomes of anxiety and depression with neuropsychological assessments of cognitive flexibility might be because neuropsychological assessments demand particular behavioural repertoires in accordance with certain contingencies, whereas self-report questionnaires rely on retrospective self-reflection (Wennerhold & Friese, 2020), which are likely tapping into different processes. Other self-report questionnaires that have been used as assessments of cognitive flexibility include the shift subscale of the Behaviour Rating Inventory of Executive Function (BRIEF; Roth, Isquith, & Gioia, 2005) and the Detail and Flexibility (DFlex) Questionnaire (Roberts, Barthel, Lopez, Tchanturia, & Treasure, 2011).

1.3.7 *The Wisconsin Card Sorting Test – an attempt to 'set the record straight'*

The WCST is a card sorting task – administered manually or via a computer – that requires participants to individually match 128 response cards to one of four stimulus cards that vary according to three dimensions: colour (red, blue, yellow, or green), form (circles, triangles, stars, or crosses) and number (one, two, three, or four). An example of the WCST is depicted in Fig 3. Rather than being explicitly informed of the correct sorting rule, participants are instead required to learn the correct sorting rule through trial and error and are provided with feedback in the form of

'correct' or 'incorrect' after each trial. In the standardised WCST manual (Heaton, Chelune, Talley, Kay, & Curtiss, 1993), the correct sorting rule changes once the participant makes ten consecutive correct responses (ten correct responses in a row = one completed category); the participant is then required to learn the new sorting rule. Each sorting rule (colour, form, and number) is repeated twice, and once the participant completes all six categories or reaches 128 trials (whichever comes first), the task is terminated.

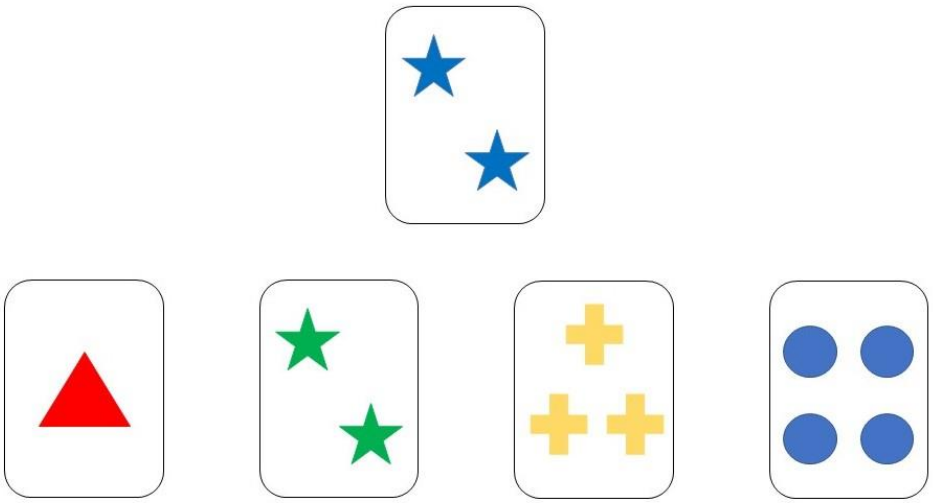


Figure 3. A depiction of the WCST. The top card (two blue stars) is the response card; the bottom four cards (one red triangle, two green stars, three yellow crosses, four blue circles) are the stimulus cards.

Card sorting tasks have a long history in psychological research (Eling, Derckx, & Maes, 2008). In particular, the WCST (Berg, 1948) has perhaps become the most widely recognised neuropsychological assessment used to assess executive functioning more broadly (Miyake, Friedman, et al., 2000; Rabin et al., 2005), and cognitive flexibility specifically, with many having used 'perseverative errors' and 'perseverative responses' as the main outcome variables of cognitive flexibility (Baker et al., 2018; Garcia-Willingham, Roach, Kasarskis, & Segerstrom, 2018; Gelonch et al., 2016; Tchanturia et al., 2012; Wiebe, Baker, Suchy, Stump, & Berg, 2018; Wollenhaupt et al., 2019). Initial investigations into what Berg (1948) referred to as *'abstraction'* and *'shifting of set'* sparked immense interest in this area of research, and 15 years later the WCST was used to study patients with frontal lobe lesions (Milner, 1963). Although the WCST has been a popular task since its inception and has been extensively used by both clinicians and researchers alike, there is no denying that the WCST is a complicated task to understand and score. Grant and

Berg (1948) were the first to provide basic scoring information of the WCST; however, it wasn't until the first manual was published by Heaton (1981) that the WCST became standardised in relation to its administration, scoring, and terminology. Despite these efforts, however, some expressed that the scoring methods outlined in the 1981 WCST manual were confusing, which gave rise to various interpretations of how to score the WCST (Flashman, Horner, & Freides, 1991; Greve, 1993). Ultimately, this urged the field to develop supplementary guidelines (Axelrod, Goldman, & Woodard, 1992; Flashman et al., 1991) until Heaton et al. (1993) published an updated WCST manual, which improved and clarified how to score perseverative responses, and provided practical examples.

Despite these advances of the WCST, considerable confusion surrounding various aspects of the task remain, particularly i) how to score the WCST according to the standardised manual (i.e., Heaton, Chelune, Talley, Kay, & Curtiss, 1993), ii) how to implement scoring principles such as the 'persevered-to' principle and the 'sandwich rule'[1], iii) the distinction between perseverative errors and perseverative responses, and iv) how to describe or define key terms of the WCST. Collectively, these issues have caused substantial confusion among users of the WCST and also present challenges when comparing and contrasting results across studies.

Identified gap: Accurate scoring of the WCST according to the standardised manual (i.e., Heaton et al., 1993), interpretation of the two main outcome variables that are often used to assess cognitive flexibility (i.e., perseverative errors and perseverative responses), and the use of key terminology of the WCST remains problematic.

Chapter 2, entitled Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility aims to clarify this gap. This chapter is published as:

*Miles, S., ***Howlett, C.A.,** Berryman, C., Nedeljkovic, M., #Moseley, G.L., & #Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, 1-9. https://doi.org/10.3758/s13428-021-01551-3

* Co-first authors – contributed equally to this manuscript.
# Co-last authors – contributed equally to this manuscript.

[1] In the WCST, the sandwich rule specifies that an ambiguous response (selecting a stimulus card that matches the response card on more than one dimension e.g., colour and form) that is preceded and proceeded by an unambiguous perseverative error and matches the perseverated-to-principle, is scored as a perseverative response. As a consequence of ambiguous responses, perseverative responses may be either correct or incorrect.

1.3.8 *The relationship between, and conflation of, self-report and neuropsychological assessments of cognitive flexibility*

Researchers and clinicians often use self-report and neuropsychological assessments of 'cognitive flexibility' interchangeably. The interchangeable use of these assessments is likely a result of the jingle fallacy – the erroneous assumption that two assessments interrogate identical constructs purely because they share similar names (Flake & Fried, 2020; Thorndike, 1913). Contrary to this assumption, however, previous research has indicated that self-report ratings and performance on neuropsychological assessments of executive function, more broadly, are poorly related (Toplak, West, & Stanovich, 2013). Specifically, a lack of association between these assessments has been found in people with mild traumatic brain injury (Spencer, Drag, Walker, & Bieliauskas, 2010), older adults with and without mild cognitive impairment (Rabin et al., 2006), and people with and without fibromyalgia (Gelonch et al., 2016). Similar findings have also been corroborated in children (Soto et al., 2020). Since executive functions are thought to comprise an array of distinct, but interconnected constructs that possibly rely on overlapping cognitive mechanisms (i.e., cognitive flexibility, inhibition, working memory etc.), assessing the relation between various self-report and neuropsychological assessments remains elusive when combining them all under the broad label of 'executive function' rather than examining such relations between assessments that are presumed to assess the *'same'* construct, i.e., 'cognitive flexibility'. While it is widely acknowledged in the field of [neuro]psychology that self-report and neuropsychological assessments of executive function are poorly associated (Toplak et al., 2013), much less is known about how self-report and neuropsychological assessments of *cognitive flexibility* specifically, relate to one another.

Attempts to clarify the relationship between self-report and neuropsychological tests of cognitive flexibility have been undertaken in various samples. In a group of non-clinical (n = 53) and clinical (comorbid anxiety and depression; n = 47) adults aged over 60 years, Johnco et al. (2014) examined the convergent validity of two self-report questionnaires of cognitive flexibility – namely the CFS and the CFI – against a range of neuropsychological tasks including the WCST, TMT, COWAT, Stroop, and the Ruff Figural Fluency Test. No significant correlations were found between the CFS, nor the CFI, and any of the neuropsychological assessments in the clinical sample. However, significant (but weak) associations between particular pairs of self-report and neuropsychological assessments were observed in the non-clinical sample. Specifically, positive correlations were found between the TMT-B and the total score of the CFI (r = .283) and the control subscale of the CFI (r = .285), respectively, as well as for the control subscale of the CFI and the

COWAT (*r* = .285) (Johnco et al., 2014). Based on these findings, Johnco (2014) argued that self-report questionnaires such as the CFS and the CFI should not be considered to be interchangeable assessments of neuropsychological assessments for assessing cognitive flexibility because they potentially assess different aspects of the construct. Another study by Lounes et al. (2011) investigated the correspondence between self-report and neuropsychological tests of cognitive flexibility in adults with and without anorexia nervosa and found no significant relationship between the CFS and the Brixton Spatial Anticipation Test in either sample. The authors suggested that these two assessments are not substitutes for one another, but could provide useful information when used together. A recent study found a small negative correlation (*r* = -.22) between the CFS and the WCST (perseverative errors) in a non-clinical sample of young adults (Grant & Cassidy, 2022). A study by Gelonch et al. (2016), revealed no significant correlations between the shift subscale of the BRIEF and the TMT (B-A), nor the WCST (perseverative errors) in patients diagnosed with fibromyalgia and healthy controls. In a study involving people with and without a lifetime diagnosis of anorexia nervosa, no correlations were detected between the various outcomes of the CFI and the TMT or WCST in the combined sample, or in either sample alone (Miles, Nedeljkovic, Sumner, & Phillipou, 2022).

The possibility that self-report and neuropsychological assessments of cognitive flexibility do not relate in adults, has support in research on adolescents (Herbrich, Kappel, Winter, & van Noort, 2019; Stedal & Dahlgren, 2015), but the available data are very limited. In a sample of patients with anorexia nervosa, Herbrich et al. (2019) found a significant moderate negative correlation between the German version of the WCST (perseverative errors) and the shift subscale of the BRIEF, indicating that a greater tendency to perseverative is associated with reduced self-reported problems in flexible shifting – a finding that contradicts what would be expected if these assessments were indeed targeting the same construct (Toplak et al., 2013). Furthermore, no significant correlations were detected between the TMT and the shift subscale of the BRIEF (Herbrich et al., 2019). Similarly, Stedal and Dahlgren (2015) found no significant correlations between the shift subscale of the BRIEF and various neuropsychological tests such as the Brixton Spatial Anticipation Test, the Stroop, TMT, and Verbal Fluency subtests from the D-KEFS that comprise part of "the Ravello Profile" – a neuropsychological test battery. These findings raise an important issue that needs to be clarified – how well do self-report and neuropsychological tests of cognitive flexibility relate to each other? To date, there has been no attempt to systematically review all the available literature to answer this question in healthy adults, let alone in clinical cohorts. Establishing the relationship between these assessments in healthy cohorts initially is an

important first step to not only strengthen the knowledge base surrounding the construct of cognitive flexibility, but to reduce the amount of 'noise' or variability that is typically associated with clinical cohorts, since many clinical conditions are confounded by comorbidities.

Identified gap: There is currently no systematic evaluation of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults.

Chapter 3, entitled Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults aims to address this question. This chapter is published as:

Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, 88(102061). https://doi.org/10.1016/j.cpr.2021.102061

Despite the importance of initially understanding whether, and to what extent, self-report and neuropsychological tests of cognitive flexibility relate in healthy adults, it is equally important to shed light on the relationship between these assessments in clinical cohorts because healthy cohorts may obscure a relationship due to ceiling effects.

Identified gap: No systematic evaluation of the relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts currently exists.

Chapter 4, entitled Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis aims to address this question. This chapter is published as:

Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology*, *36*(5), 347-372. https://doi.org/10.1037/neu0000796

The trajectory of self-report and neuropsychological assessments of 'cognitive flexibility' has, historically, remained independent. However, these two assessment approaches have begun to be considered as 'two-sides of the same coin' by clinicians and researchers – that is, they are used as equivalent and interchangeable assessments to assess the same underlying construct, and have gradually become conflated with one another. Yet, it is unclear when, why or how such conflation

has occurred, or which field has played a key role in fostering and reinforcing the conflation of these assessments. Identifying where the conflation of self-report and neuropsychological assessments of 'cognitive flexibility' has stemmed from is important so that clinicians and researchers are aware of the inadvertent consequences of doing so, and so that the field can learn from the mistakes of the past, rather than being doomed to repeat them.

Identified gap: It is unknown when, why and how self-report and neurocognitive assessments of cognitive flexibility have become conflated.

Chapter 5, entitled Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy aims to address this gap. This chapter is published as:

***Howlett, C.A.,** *Miles, S., Berryman, C., #Phillipou, A., & #Moseley, G.L. (2023). Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy. *Australian Journal of Psychology, 75*(1), 2174684. https://doi.org/10.1080/00049530.2023.2174684

* Co-first authors – contributed equally to this manuscript.
# Co-last authors – contributed equally to this manuscript.

1.4 Cognitive flexibility and persistent pain

1.4.1 *Are people with persistent pain characterised by poor cognitive flexibility?*

Interest in cognitive flexibility, indexed via neuropsychological tests, in people with persistent pain has gradually increased over the last 20 years. A substantial body of evidence from several systematic reviews has found that people with persistent pain perform significantly worse than healthy controls on several neuropsychological assessments of cognitive flexibility. Specifically, a meta-analysis involving participants with and without persistent pain (n = 3,067; persistent pain = 1,054, healthy controls = 2,013) identified small-to-moderate set-shifting impairments in people with persistent pain (Berryman et al., 2014). Corroborating this finding, another meta-analysis found small set-shifting deficits in people with fibromyalgia (n = 964) when compared to healthy controls (n = 1,025) (Bell et al., 2018). In their meta-analysis, Kessels, Aleman, Verhagen and van Luijtelaar (2000) demonstrated a moderate effect size for the difference in cognitive flexibility, as indexed by the TMT, between individuals with chronic symptomatic whiplash and healthy controls. This study also found that scores on the TMT did not improve between baseline and 6-months post-accident for those with chronic symptomatic whiplash (Kessels et al., 2000).

Combining outcomes from the TMT and the Stroop in a meta-analysis demonstrated small-to-moderate group differences, with people with persistent pain performing worse than healthy controls (Rathbone et al., 2016). These results are also supported by a recent systematic review involving people with persistent pain and healthy controls that found moderate set-shifting reductions in people with persistent pain (Bunk et al., 2019).

Although previous systematic reviews have identified a consistent small-to-moderate effect for deficits on neuropsychological assessments of cognitive flexibility, the findings from cross-sectional studies have produced conflicting findings. Neuropsychological performance was assessed by Lee et al. (2015) who found that people with complex regional pain syndrome (CRPS) made significantly more perseverative errors on the WCST than healthy controls. Similarly, Weiner, Rudy, Morrow, Slaboda, & Lieber (2006) reported that older adults with persistent low back pain performed worse than their pain-free counterparts on the TMT – part B. Corroborating this finding, Tesio et al. (2015) reported that people fibromyalgia took longer to complete the TMT – part B than healthy controls did. However, many other studies have not found such differences between people with and without persistent pain (Cherry et al., 2014; Gelonch et al., 2016; Moriarty et al., 2017; Nadar, Jasem, & Manee, 2016; Suhr, 2003; Verdejo-Garcia, Lopez-Torrecillas, Calandre, Delgado-Rodriguez, & Bechara, 2009). Inconsistent findings across studies could be a consequence of methodological limitations. In particular, some studies are underpowered with small sample sizes; possible confounds and multiple comparisons are rarely controlled for; and a diverse range of assessments are employed across studies, which makes comparing and pooling findings difficult. On the basis of these limitations, further investigation of whether people with and without persistent pain differ on neuropsychological assessments of cognitive flexibility is warranted.

While a large body of research has indicated that people with persistent pain may have reduced performance on neuropsychological assessments of cognitive flexibility, only two studies to date have examined how people with and without persistent pain differ in their ability to be flexible on self-report assessments (Baker, Gibson, Georgiou-Karistianis, Roth, & Giummarra, 2016; Gelonch et al., 2016). Baker et al. (2016) found that people with persistent pain reported reduced functioning on the shift subscale of the BRIEF than healthy controls; Gelonch et al. (2016) found that while people with fibromyalgia scored lower on the shift subscale of the BRIEF than their healthy counterparts, the group differences were no longer significant after controlling for mood

disturbances such as depression and anxiety. Gelonch et al. (2016) also correlated the shift subscale of the BRIEF with various neuropsychological tests of cognitive flexibility (e.g., scores on the TMT, Stroop and WCST) and found no relationship between any of these assessments in either cohort. This study is the only known study to i) investigate differences between people with and without persistent pain on self-report and neuropsychological assessments of cognitive flexibility concurrently and ii) examine the relationship between self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain. Thus, this is an area of research that remains limited.

Whether or not assessments of cognitive flexibility predict the presence of persistent pain seems worthy of further investigation. Of the limited data that are currently available, poor pre-operative performance on the TMT – part B has been found to predict the presence of persistent pain at 6- and 12-months post-surgery in a combined sample of patients undergoing surgery for knee osteoarthritis or breast cancer (Attal et al., 2014). However, a recent study by Vila et al. (2020) found that pre-operative performance on the TMT (B-A time) was not predictive of the presence of persistent pain at 6-months post-surgery in people undergoing knee arthroplasty and chest surgery. That the available evidence is limited, contradictory, and excludes the possible capacity of self-report questionnaires to predict the incidence of persistent pain points to a need for further investigation. One possibility that is yet to be explored is whether a combination of self-report and neuropsychological assessments of cognitive flexibility predict the presence of persistent pain. Such an enquiry may provide a foundation for more effectively tailoring interventions to individuals on the basis of those assessments.

Questions that remain to be answered:

1. Do people with persistent pain perform worse than pain-free controls on a neuropsychological assessment of cognitive flexibility?
2. Is self-reported cognitive flexibility lower in people with persistent pain than it is in pain-free controls?
3. Is there a relationship between a self-report and neuropsychological test of cognitive flexibility in people with persistent pain?
4. Is there a relationship between a self-report and neuropsychological test of cognitive flexibility in pain-free controls?
5. Is the probability of having persistent pain associated with a person's scores on a self-report and neuropsychological assessment of cognitive flexibility?

Chapter 5, entitled Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study aims to address these questions.

1.5 Thesis aim and specific research objectives

The overall aim of this thesis was to investigate the role of cognitive flexibility in persistent pain.

Specific research objectives:

1. *To provide a brief overview of how to accurately score the WCST according to the standardised manual (i.e., Heaton et al., 1993), how to interpret the two main outcome variables used to assess cognitive flexibility (i.e., perseverative errors and perseverative responses), and clarify key terminology of the WCST.*

   Chapter 2, entitled 'Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility' presents this work and was published as:

   *Miles, S., ***Howlett, C.A.,** Berryman, C., Nedeljkovic, M., #Moseley, G.L., & #Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, 1-9. https://doi.org/10.3758/s13428-021-01551-3

   * Co-first authors – contributed equally to this manuscript.
   # Co-last authors – contributed equally to this manuscript.

2. *To evaluate the current evidence concerning the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults.*

   Chapter 3, entitled 'Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults' presents this work and was published as:

   Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, 88(102061). https://doi.org/10.1016/j.cpr.2021.102061

3. *To evaluate the relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts.*

   Chapter 4, entitled 'Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis' presents this work and was published as:

   Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology,* 36(5), 347-372. https://doi.org/10.1037/neu0000796

4. *To provide a critical historical overview of when, how and why self-report and neuropsychological assessments of cognitive flexibility have become conflated.*

   Chapter 5, entitled 'Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy' presents this work and was published as:

   ***Howlett, C.A.,** *Miles, S., Berryman, C., #Phillipou, A., & #Moseley, G.L. (2023). Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy. *Australian Journal of Psychology, 75*(1), 2174684. https://doi.org/10.1080/00049530.2023.2174684

   * Co-first authors – contributed equally to this manuscript.
   # Co-last authors – contributed equally to this manuscript.

5. *To determine whether self-reported and neuropsychologically assessed cognitive flexibility is poorer in people with persistent pain than it is for healthy controls.*

   Chapter 6, entitled 'Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study' presents this work and was under review at the time of submission at *The British Journal of Pain.*

CHAPTER TWO

Chapter 2 is published as:

[*]Miles, S., [*]**Howlett, C.A.,** Berryman, C., Nedeljkovic, M., [#]Moseley, G.L., & [#]Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, 1-9. https://doi.org/10.3758/s13428-021-01551-3

* Co-first authors – contributed equally to this manuscript.

# Co-last authors – contributed equally to this manuscript.

This article was reproduced with permission from Springer Nature. A signed statement of contribution for this co-authored manuscript can be found in Appendix F.

Behavior Research Methods
https://doi.org/10.3758/s13428-021-01551-3

Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility

Stephanie Miles[1] · Caitlin A. Howlett[2] · Carolyn Berryman[3] · Maja Nedeljkovic[1] · G. Lorimer Moseley[2] · Andrea Phillipou[1,4,5,6]

Accepted: 22 January 2021
© The Psychonomic Society, Inc. 2021

Abstract

The Wisconsin Card Sorting Test (WCST) is a popular neurocognitive task used to assess cognitive flexibility, and aspects of executive functioning more broadly, in research and clinical practice. Despite its widespread use and the development of an updated WCST manual in 1993, confusion remains in the literature about how to score the WCST, and importantly, how to interpret the outcome variables as indicators of cognitive flexibility. This critical review provides an overview of the changes in the WCST, how existing scoring methods of the task differ, the key terminology and how these relate to the assessment of cognitive flexibility, and issues with the use of the WCST across the literature. In particular, this review focuses on the confusion between the terms 'perseverative responses' and 'perseverative errors' and the inconsistent scoring of these variables. To our knowledge, this critical review is the first of its kind to focus on the inherent issues surrounding the WCST when used as an assessment of cognitive flexibility. We provide recommendations to overcome these and other issues when using the WCST in future research and clinical practice.

Keywords Cognitive flexibility · Set-shifting · Executive function · WCST · Neurocognitive measures · Neurocognitive assessment

Cognitive flexibility refers to the mental ability that enables us to effectively adjust to changing task and/or environmental demands (Deák, 2003; Scott, 1962) and is thought to arise from the interaction between higher-order executive functions (Dajani & Uddin, 2015). Further, cognitive flexibility is a critical cognitive function that allows an individual to switch their cognitive strategies, consider two or more aspects of an object, idea, or complex situation simultaneously, and appropriately adapt behavioural strategies (Bilgin, 2009; Dennis & Vander Wal, 2010; Diamond, 2013). There are many strategies by which researchers and clinicians assess cognitive flexibility, one of which is the use of neurocognitive tests.

The most commonly used neurocognitive test for assessing cognitive flexibility is the Wisconsin Card Sorting Test (WCST; Berg, 1948; Johnco, Wuthrich, & Rapee, 2014; Tchanturia et al., 2012). Originally popular as an assessment of frontal lobe damage (Barceló & Knight, 2002; Milner, 1963), by 2005, the WCST was the seventh most frequent test implemented by clinical neuropsychologists (Rabin, Barr, & Burton, 2005). The WCST yields upwards of seven variables. 'Perseverative responses' and 'perseverative errors' are the most commonly used to assess cognitive flexibility (Baker, Georgiou-Karistianis, et al., 2018a; Baker, Gibson, Georgiou-Karistianis, & Giummarra, 2018b; Dickson, Ciesla, & Zelic, 2017; Garcia-Willingham, Roach, Kasarskis, & Segerstrom, 2018; Gelonch, Garolera, Valls, Rosselló, & Pifarré, 2016; Wollenhaupt et al., 2019), but

Stephanie Miles and Caitlin A. Howlett are Co first authors

G. Lorimer Moseley and Andrea Phillipou are Co last authors

✉ Stephanie Miles
smiles@swin.edu.au

1 Centre for Mental Health, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia

2 Innovation, Implementation & Clinical Translation (IIMPACT) in Health, Allied Health and Human Performance, University of South Australia, Adelaide, Australia

3 Brain Stimulation, Imaging and Cognition Research Group, School of Medicine, The University of Adelaide, Adelaide, Australia

4 Department of Mental Health, St Vincent's Hospital, Melbourne, Australia

5 Department of Psychiatry, The University of Melbourne, Melbourne, Australia

6 Department of Mental Health, Austin Health, Melbourne, Australia

'number of categories completed', 'failure-to-maintain-set', 'trials to complete the first category' and 'non-perseverative errors' have also been used (e.g. Abbate-Daga, Buzzichelli, Marzola, Amianto, & Fassino, 2014; Aloi et al., 2015; Bischoff-Grethe et al., 2013; Dickson et al., 2017; Gelonch et al., 2016; Tchanturia et al., 2012; Wollenhaupt et al., 2019; Zmigrod, Rentfrow, & Robbins, 2018). This has led to criticism of the WCST for having too many outcome variables (Figueroa & Youmans, 2013), that are not clearly linked to particular cognitive domains (Greve, Stickle, Love, Bianchini, & Stanford, 2005). Moreover, various definitions and interchanging of variables, inconsistency in how to obtain the variables, and unclear, incomplete, or misleading reporting clouds the field. Here we highlight several critical problems with the current use of the WCST and recommend solutions for implementation in research and clinical practice.

An overview of the Wisconsin Card Sorting Test

The WCST is a card matching task. For each trial, a response card is placed above four multidimensional stimulus cards (see Fig. 1 for an example of the WCST). The cards presented in the task vary on three dimensions —colour (red, blue, yellow, green), form (circles, triangles, stars, crosses), and number (one, two, three, four). For each 'trial', participants 'match' the response card to one of the four stimulus cards, without specific instruction by the administrator. The sorting rule is the dimension on which the card needs to be correctly matched, and the participant identifies the sorting rule through a process of trial and error. For example, a response card with two blue triangles can be matched according to colour (blue), form (triangle), or number (two). After each response, the participant receives feedback (i.e. 'correct' or 'incorrect') that can be used to establish the correct sorting rule. Typically, the sorting rule changes without warning after ten correct responses in a row (this is called 'completing a category') and the participant must 'start again' to establish the new sorting rule for the following category. In the standard administration of the task, the order of the sorting rule is colour, form, number, colour, form, number (Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The WCST terminates when either (i) all six categories are completed, or (ii) 128 trials are completed. The WCST was first developed as a manually administered task using physical cards; however, it has since been adapted to a computerised format (Heaton & PAR Staff, 2008), and is widely used in both physical and electronic forms.

Changes in the WCST since its development

The WCST was first developed by Berg (1948) to assess flexibility in thinking, but has since seen some changes (see Eling, Derckx, & Maes, 2008, for a full historical and conceptual review). Briefly, the original WCST (Berg, 1948) required participants to make five correct responses in a row before the sorting rule changed (it is now ten), and nine categories were completed (it is now six). The original WCST consisted of 60 stimulus cards (Berg, 1948); Grant and Berg (1948) modified the task to include 64 stimulus cards, which were repeated if the participant used all 64 cards before completing the task (i.e. completing all categories). The original WCST had no limits on the number of trials taken to complete the task (Berg, 1948); there are now a maximum of 128 trials (Heaton et al., 1993); however, shorter and perhaps more practical versions of the WCST have been developed and are commonly used in clinical settings (i.e. the WCST-64 card version; Greve, 2001). Berg (1948) considered only the average number of errors made and the number of categories

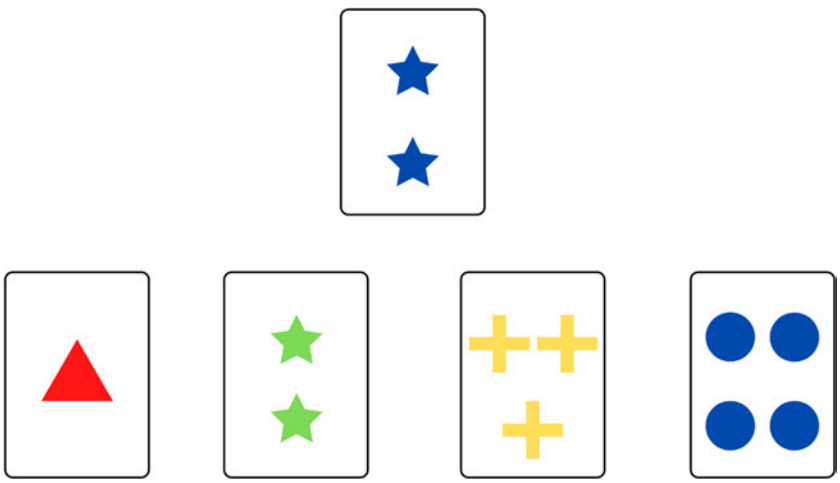


Fig. 1 An example of the WCST display. The figure depicts a response card with two blue stars. The stimulus cards are from left to right; one red triangle, two green stars, three yellow crosses, and four blue circles. In the manual administration, the four stimulus cards would be laid on a table in front of the participant, and the participant is handed a deck of multidimensional response cards to sort

completed during the WCST; Grant and Berg (1948) added perseverative and non-perseverative errors as key outcome variables for the WCST and provided basic scoring information for the task.

Despite the test's development over time, its popularity, and widespread use, little to no standardisation of the administration, scoring, and terminology existed until Heaton, Chelune, Talley, Kay, and Curtiss (1981) published a standardised manual. Even then, the administration and scoring rules remained unclear to many (Flashman, Horner, & Freides, 1991; Greve, 1993), and supplementary, more transparent scoring guidelines were developed (e.g. Axelrod, Goldman, & Woodard, 1992; Flashman et al., 1991). An updated WCST manual was published in 1993 (Heaton et al., 1993), providing clarification for how to score perseverative responses, one of the key variables used today to assess cognitive flexibility.

Inconsistent terminology and unclear definitions

There are inconsistencies in the WCST terminology used throughout the literature, and the definitions used to operationalise these terms vary between studies. These inconsistencies have caused significant confusion for both researchers and clinicians who use the WCST, and have led to further discrepancies, bespoke scoring approaches, and a lack of transparency in reporting administration and scoring. Many sources of confusion exist. For example, Heaton et al. (1993) used '*dimension*' to refer to the characteristics of the cards by which they can be sorted (i.e. colour, form, number, or other); WCST studies have variably used *attribute*, *characteristic*, *criterion*, *rule*, *category*, *or principle* in lieu of dimension. The term '*category*' refers to the discrete sections of the WCST (Grant & Berg, 1948; Heaton et al., 1993). There are six categories, which follow the order colour, form, number, colour, form, number. The category determines the correct sorting rule. Given that the dimensions are repeated twice, categories are sometimes referred to in terms of their numerical order (e.g. first category).

Confusion and problems in scoring the WCST

The 'perseverated-to' principle

The 'perseverated-to' principle, a key scoring principle of the WCST, remains a poorly understood concept. The perseverated-to principle was formally conceptualised by Heaton et al. (1993) after supplementary scoring guidelines for the WCST were developed by Flashman et al. (1991). The perseverated-to principle was described by Heaton et al. (1993) in conjunction with a definition of perseveration, which may have caused confusion between the perseverated-to principle and the meaning of perseverative responses. In an attempt to clarify the key scoring terms of the WCST, we have provided definitions in Table 1 that are in line with the standardised manual (Heaton et al., 1993). The perseverated-to principle is established in the first category of the WCST after the first unambiguous incorrect response. Subsequent responses that match the perseverated-to principle are scored as perseverative. However, an ambiguous response must also meet the criteria for the 'sandwich rule', which specifies that an ambiguous response preceded and followed by an unambiguous response must match the dimension of the unambiguous response, thus demonstrating a perseverative response pattern.[1] Following the completion of a category (i.e. when ten correct responses are made in a row), the previously *correct sorting rule* becomes the *perseverated to principle* in the next category (Heaton et al., 1993). For example, if the correct sorting rule in the previous category was colour and the correct sorting rule for the current category is form, colour becomes the perseverated-to principle (see Fig. 2, column (a) for an example). According to Heaton et al. (1993), the perseverated-to principle can also change within a category. Following from the previous example, if the participant makes three sequential unambiguous errors according to number within the form category (during which colour is the initial perseverated-to-principle), then the dimension 'number' becomes the new perseverated-to principle (see Fig. 2, column (a) for an example). It is important to note that in this situation the new perseverated-to principle does not come into effect until the second unambiguous error, and thus the second unambiguous error (not the first) is the first response to be scored as perseverative.

Perseverative responses and perseverative errors: Two key variables to assess cognitive flexibility

Perseverative responses are persistent responses made by a participant on the basis of an incorrect (previous or novel) stimulus dimension (Heaton et al., 1993; McCallum, 2017). For example, a participant may attempt to sort a card based on its form, receive feedback that this response is incorrect, and yet continue to make the same mistake on the following card. Typically, perseverative responses are incorrect (i.e. the response does not match the correct sorting rule), and are termed perseverative errors (Grant & Berg, 1948). However, a perseverative response may also be correct because of the

[1] The sandwich rule was first conceptualised by Grant & Berg (1948), fully developed by Flashman et al. (1991) and later incorporated into the updated WCST manual by Heaton et al. (1993)

Table 1 Key WCST scoring terminology derived from the revised WCST manual (Heaton et al., 1993)

Correct the response given matches the correct sorting rule	**Incorrect** the response given does not match the correct sorting rule
Correct sorting rule the dimension that participants are meant to be sorting the cards by, i.e. the rule that is in effect for that category	**Perseverated to principle** the incorrect sorting dimension which a participant is repeatedly responding to (e.g. form, when the correct sorting rule is colour). The perseverated to principle applies to only one dimension at a time (i.e. colour or form or number)
Perseverative responses when the dimension the participant is sorting by matches the perseverated to principle, i.e. the participant continually responds using an incorrect dimension even after receiving feedback that this response is incorrect. Participants cannot perseverate on the *other* dimension	**Non perseverative responses** when the dimension the participant is using to sort does not match the perseverated to principle
Ambiguous when the stimulus card chosen matches the response card on two or more dimensions (e.g. colour and form). By definition, responses that match the *other* dimension are ambiguous. Figure 1 provides an illustration of a potentially ambiguous response. The response card matches the second stimulus card on both form and number (two stars)	**Unambiguous** when the stimulus card chosen matches the response card on only one dimension (e.g. colour). Figure 1 provides an illustration of a potentially unambiguous response, in that the response card matches the last stimulus card on colour only (blue)

potential for ambiguous responses and the above-mentioned sandwich rule. On some trials of the WCST, the chosen stimulus card will match the response card on multiple dimensions (e.g. the response card 'four green circles' matches the stimulus card 'four blue circles' on both number and form). These sorts of responses are considered ambiguous because the administrator cannot be certain of which dimension the participant is using to sort the card. Typically, an inspection of the unambiguous responses before and after an ambiguous response can indicate the sorting rule that a participant is using. According to the sandwich rule, an ambiguous response that follows and is followed by an unambiguous perseverative error, and matches the perseverated-to principle, is scored as a perseverative response (see Fig. 2, column (b) for an example of this). Thus, even in cases wherein the ambiguous response matches the correct sorting rule, the response is scored as perseverative, i.e. it is a perseverative response, but not a perseverative error. To summarise: all perseverative errors are perseverative responses, but not all perseverative responses are perseverative errors (see Fig. 3). The terms have been

(a)

Number sequentially correct	Response	Perseverative response
Correct sorting rule: colour, followed by form		
10	**C** F N O	
	C F N O ptp	p
	C F N O	p
	C F **N** O	
	C F **N** O ptp	p
	C F **N** O	p
	C F **N** O	p

(b)

Number sequentially correct	Response	Perseverative response
Correct sorting rule: number, followed by form		
10	C F **N** O	
	C F **N** O ptp	p
	C F **N** O	p
1	C **F** **N** O	p
	C F **N** O	p
	C F **N** O	p
1	C **F** N O	

(c)

Number sequentially correct	Response	Perseverative response
Correct sorting rule: colour, followed by form		
8	**C** F N O	
9	**C** F **N** O	
10	**C** F N O	
	C F N O ptp	p
	C F N O	p
	C F **N** O	
1	C **F** N O	

Fig. 2 Examples of various scoring scenarios for the WCST according to the Heaton et al. (1993) scoring manual. *Note*: Each main column comprises three scoring columns Number sequentially correct, Response, and Perseverative response. Each column is divided by a line at the point of category and rule change. Column (a) provides an illustration of the perseverated to principle after a category change and how the perseverated to principle can change within a category after three unambiguous responses; column (b) provides an illustration of how to score perseverative responses according to the sandwich rule when there is an ambiguous response; and column (c) provides an illustration of how to score perseverative errors. The number in the correct column signifies the number of correct sequential responses; a bolded and underlined letter indicates which dimensions the participant's response matches; C = colour; F = form; N = number; O = other (i.e. the response did not match on any of the three key dimensions); p = perseverative response; ptp = indicates the establishment of the perseverated to principle

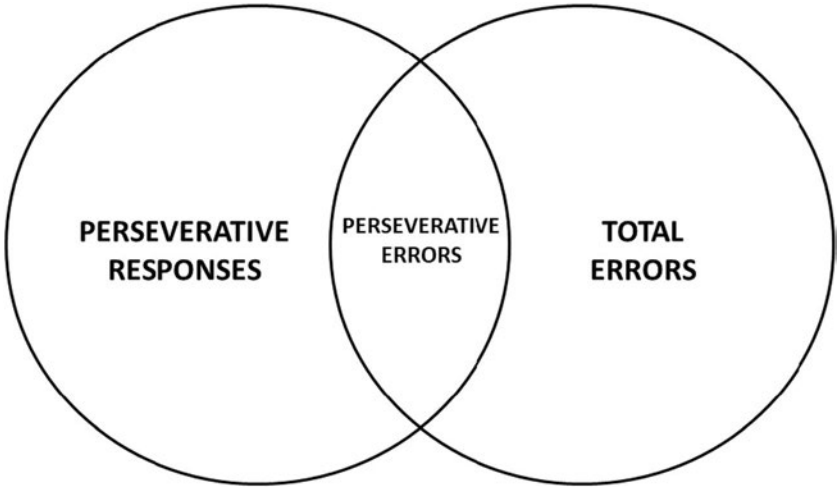


Fig. 3 A Venn diagram illustrating the relationship between perseverative responses and errors made on the WCST

sometimes used interchangeably (e.g. Gelonch et al., 2016; Øverås, Kapstad, Brunborg, Landrø, & Lask, 2015; Strauss, Sherman, & Spreen, 2006), illustrating the potential for erroneous results for both variables. This problem even infiltrates the revised WCST manual (Heaton et al., 1993), which uses the terms interchangeably and fails to highlight whether perseverative responses, perseverative errors, or both, should be reported.

Grant and Berg (1948) and Heaton et al. (1993) have different methods for scoring perseverative responses and perseverative errors. Grant and Berg (1948) specified that perseverative responses occurred after a rule change and were responses that matched the sorting rule of the previous category. Yet, the more contemporary definition provided by the revised WCST manual specifies that perseverative responses are those that match the perseverated-to principle (Heaton et al., 1993). Thus, a perseverative response (or perseverative error) can occur at any point in the task, including before a rule change and during the first category of the task (Heaton et al., 1993). Problematically, both scoring methods continue to be used in current research, which is likely to create difficulties in comparing results across studies.

When describing the number of categories achieved on the WCST, Strauss et al. (2006) explain that 'scores can range from 0 for the subject who never gets the idea at all to 6' (p. 528–529). Paradoxically, a participant who never learns the correct sorting rule (i.e. they fail to achieve ten correct responses in a row), and thus never advances from the first category, will receive a score of zero perseverative errors under the Grant and Berg (1948) scoring method. Considered to reflect high cognitive flexibility, a score of zero perseverative errors in this situation would be not only imprecise but contradictory. In these instances, the Grant and Berg (1948) scoring method provides an erroneous impression of cognitive flexibility. Typically, in instances where the participant has failed to complete the first category and the Grant and Berg (1948) method has been used, if the participant's responses are examined in greater detail, it becomes evident that the participant has perseverated within the first failed category. For example, the participant may have sorted according to an incorrect dimension several times in a row before testing another dimension, thus demonstrating cognitive rigidity. Consequently, in these cases, the Heaton et al. (1993) scoring method may be more appropriate to capture all instances of perseverative responses. Although the number of categories completed may not be of importance when considering other WCST variables (e.g. failure-to-maintain-set), it is evident that the number of categories completed in conjunction with the scoring method used have an influence on the number or percentage of perseverative responses scored.

Perhaps in response to the problems outlined, some authors implement their own scoring method. The implementation of different scoring techniques in some papers has led to discrepancies in the total number/percentage of perseverative responses scored. For example, according to the 1993 WCST manual, if the first response after the completion of a category (i.e. the first trial in which a new sorting rule is in effect) is an unambiguous error that matches the previous sorting rule, it is marked as a perseverative error (see Fig. 2, column (c); Heaton et al., 1993). However, in Channon (1996), these responses were not scored as perseverative errors '[because] the sorting principle changed without warning and they had no means of knowing this in advance' (p. 109). In both examples, the sorting rule changes without warning at the end of a category. The two scoring techniques represent differences in the approach to, and understanding of, perseverative responses. Indeed, Channon's (1996) method of scoring would afford up to six fewer perseverative responses than Heaton et al. (1993), making studies using these different methods incomparable. There are many cases of studies implementing their own scoring techniques for the WCST or failing to specify the

scoring method at all (e.g. Perpiñá, Segura, & Sanchez-Reales, 2017; Van Eylen et al., 2011). The inconsistencies in scoring methods present challenges when comparing results and generalising findings.

Using the WCST to assess cognitive flexibility

Although the WCST is widely accepted as an assessment of cognitive flexibility (Cragg & Chevalier, 2012; Dennis & Vander Wal, 2010; Figueroa & Youmans, 2013), neurocognitive tasks inevitably assess other executive functions (Buchsbaum, Greer, Chang, & Berman, 2005; Miyake, Emerson, & Friedman, 2000). Importantly, the WCST is a complex task that requires the use of multiple executive functions including attention, memory, and implicit learning (Buchsbaum et al., 2005; Cepeda, Kramer, & Gonzalez de Sather, 2001; Friederich & Herzog, 2011; Wu et al., 2014). Consequently, when the WCST is used to assess cognitive flexibility, the influence of other executive functions should be considered and accounted for where possible.

Typically, perseverative responses and/or perseverative errors are presented as indicators of cognitive flexibility (e.g. Baker, Georgiou-Karistianis, et al., 2018a; Baker, Gibson, et al., 2018b; Dickson et al., 2017; Garcia-Willingham et al., 2018; Gelonch et al., 2016; Wollenhaupt et al., 2019). However, some research has also used other variables (e.g. the number of categories completed, the number of trials taken to complete the first category, non-perseverative errors, and failure-to-maintain-set) as indicators of cognitive flexibility (e.g. Abbate-Daga et al., 2014; Aloi et al., 2015; Bischoff-Grethe et al., 2013; Dickson et al., 2017; Gelonch et al., 2016; Tchanturia et al., 2012; Wollenhaupt et al., 2019; Zmigrod et al., 2018). The extent to which some of these variables assess cognitive flexibility remains under debate; we recommend the reader refer to Figueroa and Youmans (2013) for discussion of the failure-to-maintain-set variable and its appropriateness as a measure of cognitive flexibility or distractibility. Further, despite the consensus that perseverative responses and/or perseverative errors are indicative of cognitive flexibility, there is no empirical evidence that conclusively verifies that these variables assess this construct. Rather, there is a general acceptance that a pattern of repetitive incorrect responding suggests rigidity and an inability to adapt to change. Comparing the WCST variables, particularly perseverative responses and/or perseverative errors, to other accepted neurocognitive tasks commonly used to assess cognitive flexibility (e.g. the Trail Making Test (TMT); Reitan, 1958) would be a first step in establishing the validity of these outcomes as markers of cognitive flexibility. Previous studies have identified a moderate to non-existent relationship between the cognitive flexibility outcomes of the WCST and the TMT (Chaytor, Schmitter-Edgecombe, & Burr, 2006; Herbrich, Kappel, Winter, & van Noort, 2019; Kortte, Horner, & Windham, 2002; O'donnell, Macgregor, Dabrowski, Oestreicher, & Romero, 1994; Pignatti & Bernasconi, 2013; Van Autreve, De Baene, Baeken, van Heeringen, & Vervaet, 2013). However, it is noteworthy that these studies have correlated diverse outcomes of these tests, and some of the reported variables may not be appropriate for assessing cognitive flexibility (i.e. TMT – Part B; Vall & Wade, 2015). Establishing consensus on which WCST and TMT variables to use when assessing cognitive flexibility will enable the field to compare findings across studies to elucidate the construct validity of such neurocognitive tests. In addition, leveraging advances in the study of cellular, circuit-level, and whole-brain imaging to uncover the neural correlates of cognitive flexibility is necessary to move the field forward (Lie, Specht, Marshall, & Fink, 2006; Specht, Lie, Shah, & Fink, 2009; Yuan & Raz, 2014).

Recommendations and conclusions

Establishing consensus on how to best define the WCST key terms and variables should be made a priority to promote the standardisation of assessments and comparability of results across studies. We recommended that perseverative responses and perseverative errors be defined and reported separately. Given that the Heaton et al. (1993) method of scoring is perhaps better able to assess perseverative responses in the first category of the WCST, we recommend that the Heaton et al. (1993) method be used to score the WCST. However, we acknowledge the complexity of this scoring system and recognise that training may be required before administrators are confident in scoring and interpreting any WCST data using the Heaton et al. (1993) method. Automated scoring that is facilitated by a computer program may be the best choice for a novice administrator. The use of a computerised version of the WCST (ideally the Heaton and PAR Staff (2008) program) reduces the opportunity for human error and misinterpretation of the scoring instructions described by Heaton et al. (1993). A non-commercial open-source version could be utilised in research and in clinical practice. In this instance, we recommend that the scoring code conform to the scoring methods outlined in the Heaton et al. (1993) manual. In instances where other computerised versions are implemented, manually inspecting the raw data to ensure that there are no atypical summary scores is an appropriate precaution to take.

To reduce the complexity of scoring the WCST, alternate versions of the test have been developed with all ambiguous cards removed (e.g. the Modified Card Sorting Test; Nelson, 1976). However, more research is needed to examine the extent to which the scores from the original WCST (e.g. perseverative errors and perseverative responses) relate to the scores

Table 2 A checklist of recommendations for using and reporting the WCST

Topic	Checklist item
Points of consideration before using the WCST as an assessment of cognitive flexibility	
Administration	Decide on which version of the WCST will be used, i.e. manual administration or computerised version.
Scoring	Decide on which scoring method will be used. Where possible, we recommend that that Heaton et al. (1993) administration and scoring be implemented. Any adaptation of the WCST should be appropriately justified and empirically supported.
Outcome variables	Establish primary outcome variables, i.e. perseverative responses (sum and/or percent age of total trials) and perseverative errors (sum and/or percentage of total trials).
Items to report in a manuscript	
Administration and scoring	Explicitly state which scoring method is used. Where possible, we recommend that that Heaton et al. (1993) administration and scoring be implemented. Any adaptation of the WCST should be appropriately justified and empirically supported.
	Explicitly state whether the task is a manual administration or computerised version.
Defined terms	*Dimension*: the characteristics of the cards by which they can be sorted (i.e. colour, form, number, or other). Avoid using the following terms: attribute, characteristic, criterion, rule, category, or principle.
	Category: refers to the discrete sections of the WCST.
	Perseverative response: persistent responses made by a participant on the basis of an incorrect (previous or novel) stimulus dimension. Perseverative responses can be correct or incorrect.
	Perseverative error: perseverative responses that are incorrect.
Variables for assessing cognitive flexibility	Report appropriate variables when assessing cognitive flexibility, i.e. perseverative responses (sum and/or percentage of total trials) and perseverative errors (sum and/or percentage of total trials).
	Avoid reporting WCST variables which may assess other executive functions as indicators of cognitive flexibility (e.g. failure to maintain set, non perseverative er rors). These WCST variables can be reported if required (e.g. when assessing other domains of executive functioning), providing they are not described as indicators of cognitive flexibility.

of the modified versions. The reduction in the total number of cards presented due to the removal of ambiguous cards creates fewer opportunities for a participant to initiate cognitive flexibility. Relatedly, the maximum number of possible perseverative responses in modified versions of the WCST is lower than in the original WCST. Hence, raw scores on the modified WCST may not be directly comparable with raw scores on the original WCST, potentially leading to further confusion in the literature, concerns surrounding validity, and problems with study comparability. We recommend that future research investigate the validity of the different outcomes of the modified WCST to provide clarification on whether removing ambiguous cards from the WCST is an appropriate step to improve scoring simplicity.

In conclusion, the WCST is undoubtedly a popular neurocognitive task that has been widely used by both researchers and clinicians since its inception. Despite the wide implementation of the WCST within the field of neuropsychology, this task is not without its limitations. The inconsistent scoring of the WCST is a major source of confusion for users and creates challenges in interpreting and comparing findings. There has also been a lack of consensus in the key WCST terminology and its corresponding definitions. Further, the outcome variables for assessing cognitive flexibility vary among studies. We recommend that users of the WCST follow the recommendations described above and presented in Table 2. Specifically, it is fundamental that authors who are considering using the WCST are transparent about the format of the task, cite the chosen scoring method, and report both perseverative responses and perseverative errors when using this task as an index to assess cognitive flexibility so as to better capture this latent construct.

Funding SM and CAH are supported by Research Training Program (RTP) scholarships (Australian Government, Department of Education, Skills and Employment). CB is supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship [ID 1127155]. GLM is supported by a NHMRC Leadership Investigator Grant [ID 1178444]. AP is supported by a NHMRC Project Grant [ID 1159953]. These funding bodies had no role in the analysis or interpre tation of the data, writing of the manuscript, or the decision to submit the paper for publication.

Declarations

Conflicting Interests GLM has received support from: ConnectHealth UK, Seqirus, Kaiser Permanente, Workers' Compensation Boards in Australia, Europe and North America, AIA Australia, the International Olympic Committee, Port Adelaide Football Club, Arsenal Football Club. Professional and scientific bodies have reimbursed him for travel costs related to presentation of research on pain at scientific conferences/ symposia. He has received speaker fees for lectures on pain and rehabil itation. He receives book royalties from NOIgroup publications, Dancing Giraffe Press & OPTP for books on pain and rehabilitation. CB has received support from Workers' Compensation Boards in Australia. All remaining authors (SM, CAH, MN, and AP) declare no potential conflicts of interest (personal or financial) with respect to the research, authorship, and/or publication of this manuscript.

References

Abbate Daga, G., Buzzichelli, S , Marzola, E., Amianto, F., & Fassino, S. (2014). Clinical investigation of set shifting subtypes in anorexia nervosa. *Psychiatry Research*, *219*(3), 592 597. https://doi.org/10.1016/j.psychres 2014.06.024

Aloi, M., Rania, M , Caroleo, M., Bruni, A., Palmieri, A., Cauteruccio, M. A., … Segura García, C. (2015). Decision making, central co herence and set shifting: A comparison between binge eating disor der, anorexia nervosa and healthy controls. *BMC Psychiatry*, *15*(1), 1 10. https://doi.org/10.1186/s12888 015 0395 z

Axelrod, B. N., Goldman, R. S., & Woodard, J. L. (1992). Interrater reliability in scoring the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *6*(2), 143 155.

Baker, K. S., Georgiou Karistianis, N., Lampit, A., Valenzuela, M., Gibson, S. J , & Giummarra, M. J. (2018a). Computerised training improves cognitive performance in chronic pain: A participant blinded randomised active controlled trial with remote supervision. *Pain*, *159*(4), 644 655. https://doi.org/10.1097/j.pain.0000000000001150

Baker, K. S., Gibson, S. J., Georgiou Karistianis, N., & Giummarra, M. J. (2018b). Relationship between self reported cognitive difficulties, objective neuropsychological test performance and psychological distress in chronic pain. *European Journal of Pain*, *22*(3), 601 613. https://doi.org/10.1002/ejp.1151

Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*, *40*(3), 349 356. https://doi.org/10.1016/S0028 3932(01)00110 5

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, *39*, 15 22.

Bilgin, M. (2009). Developing a cognitive flexibility scale: Validity and reliability studies. *Social Behavior and Personality: An International Journal*, *37*(3), 343 353. https://doi.org/10.2224/sbp.2009.37.3.343

Bischoff Grethe, A., McCurdy, D., Grenesko Stevens, E., Irvine, L. E., Wagner, A., Yau, W. Y. W., … Kaye, W. H. (2013). Altered brain response to reward and punishment in adolescents with anorexia nervosa. *Psychiatry Research Neuroimaging*, *214*(3), 331 340. https://doi.org/10.1016/j.pscychresns.2013.07.004

Buchsbaum, B. R., Greer, S., Chang, W. L., & Berman, K. F. (2005). Meta analysis of neuroimaging studies of the Wisconsin Card Sorting Task and component processes. *Human Brain Mapping*, *25*(1), 35 45. https://doi.org/10.1002/hbm.20128

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. M. (2001). Changes in executive control across the life span: Examination of task switching performance. *Developmental Psychology*, *37*(5), 715 730. https://doi.org/10.1037/0012 1649.37.5.715

Channon, S. (1996). Executive dysfunction in depression: The Wisconsin Card Sorting Test. *Journal of Affective Disorders*, *39*(2), 107 114. https://doi.org/10.1016/0165 0327(96)00027 4

Chaytor, N., Schmitter Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*(3), 217 227. https://doi.org/10.1016/j.acn.2005.12.002

Cragg, L., & Chevalier, N. (2012). The processes underlying flexibility in childhood. *Quarterly Journal of Experimental Psychology*, *65*(2), 209 232. https://doi.org/10.1080/17470210903204618

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571 578. https://doi.org/10.1016/j.tins.2015.07.003

Deák, G. O. (2003). The development of cognitive flexibility and lan guage abilities. *Advances in Child Development and Behavior*, *31*, 271 327. https://doi.org/10.1016/S0065 2407(03)31007 9

Dennis, J. P., & Vander Wal, J. S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research*, *34*(3), 241 253. https://doi.org/10.1007/s10608 009 9276 4

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135 168. https://doi.org/10.1146/annurev psych 113011 143750

Dickson, K. S., Ciesla, J. A., & Zelic, K. (2017). The role of executive functioning in adolescent rumination and depression. *Cognitive Therapy and Research*, *41*(1), 62 72. https://doi.org/10.1007/s10608 016 9802 0

Eling, P., Derckx, K., & Maes, R. (2008). On the historical and concep tual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, *67*(3), 247 253. https://doi.org/10.1016/j.bandc.2008.01.006

Figueroa, I. J., & Youmans, R. J. (2013). Failure to maintain set: A measure of distractibility or cognitive flexibility? *Proceedings of the Human Factors and Ergonomics Society*, *57*(1), 828 832. https://doi.org/10.1177/1541931213571180

Flashman, L. A., Horner, M. D., & Freides, D. (1991). Note on scoring perseveration on the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *5*(2), 190 194.

Friederich, H. C., & Herzog, W. (2011). Cognitive behavioural flexibil ity in anorexia nervosa. In R. A. H. Adan & W. H. Kaye (Eds.), *Behavioral neurobiology of eating disorders (Current topics in be havioral neurosciences)* (6th ed., pp. 111 123). Berlin: Springer Verlag. https://doi.org/10.1007/7854_2010_83

Garcia Willingham, N. E., Roach, A. R., Kasarskis, E. J., & Segerstrom, S. C. (2018). Self regulation and executive functioning as related to survival in motor neuron disease: Preliminary findings. *Psychosomatic Medicine*, *80*(7), 665 672. https://doi.org/10.1097/PSY.0000000000000602

Gelonch, O., Garolera, M., Valls, J , Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and ob jective measures. *Comprehensive Psychiatry*, *66*, 113 122. https://doi.org/10.1016/j.comppsych.2016.01.002

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl type card sorting problem. *Journal of Experimental Psychology*, *38*(4), 404 411.

Greve, K. W. (1993). Can perseverative responses on the Wisconsin Card Sorting Test be scored accurately? *Archives of Clinical Neuropsychology*, *8*(6), 511 517. https://doi.org/10.1016/0887 6177(93)90051 2

Greve, K. W. (2001). The WCST 64: A standardized short form of the Wisconsin Card Sorting Test. *Clinical Neuropsychologist*, *15*(2), 228 234. https://doi.org/10.1076/clin.15.2.228.1901

Greve, K. W., Stickle, T. R., Love, J. M., Bianchini, K. J., & Stanford, M. S. (2005). Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study. *Archives of Clinical Neuropsychology*, *20*(3), 355 364. https://doi.org/10.1016/j.acn.2004.09.004

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1981). *Wisconsin Card Sorting Test: Manual* (1st ed.). Odessa: Psychological Assessment Resources.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Odessa: Psychological Assessment Resources.

Heaton, R. K., & PAR Staff. (2008). Wisconsin Card Sorting Test: Computer Version 4 Research Edition. PAR.

Herbrich, L. R., Kappel, V., Winter, S. M., & van Noort, B. M. (2019). Executive functioning in adolescent anorexia nervosa: Neuropsychology versus self and parental report. *Child Neuropsychology*, *25*(6), 816 835. https://doi.org/10.1080/09297049.2018.1536200

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). Reliability and validity of two self report measures of cognitive flexibility. *Psychological Assessment*, *26*(4), 1381 1387. https://doi.org/10.1037/a0038009

Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The Trail Making Test, Part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, *9*(2), 106 109. https://doi.org/10.1207/S15324826AN0902_5

Lie, C. H., Specht, K., Marshall, J. C., & Fink, G. R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. *NeuroImage*, *30*(3), 1038 1049. https://doi.org/10.1016/j.neuroimage.2005.10.031

McCallum, R. S. (2017). *Handbook of nonverbal assessment*. https://doi.org/10.1007/978 3 319 50604 3

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology*, *9*(1), 90 100. https://doi.org/10.29085/9781856048767.060

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *21*(2), 169 183.

Nelson, H. E. (1976). A Modified Card Sorting Test sensitive to frontal lobe defects. *Cortex*, *12*(4), 313 324. https://doi.org/10.1016/S0010 9452(76)80035 4

O'Donnell, J. P., Macgregor, L. A., Dabrowski, J. J., Oestreicher, J. M., & Romero, J. J. (1994). Construct validity of neuropsychological tests of conceptual and attentional abilities. *Journal of Clinical Psychology*, *50*(4), 596 600. https://doi.org/10.1002/1097 4679(199407)50:4<596::AID JCLP2270500416>3.0.CO;2 S

Øverås, M., Kapstad, H., Brunborg, C., Landrø, N. I., & Lask, B. (2015). Are poor set shifting abilities associated with a higher frequency of body checking in anorexia nervosa? *Journal of Eating Disorders*, *3*(1), 1 8. https://doi.org/10.1186/s40337 015 0053 3

Perpiñá, C., Segura, M., & Sanchez Reales, S. (2017). Cognitive flexibility and decision making in eating disorders and obesity. *Eating and Weight Disorders*, *22*(3), 435 444. https://doi.org/10.1007/s40519 016 0331 3

Pignatti, R., & Bernasconi, V. (2013). Personality, clinical features, and test instructions can affect executive functions in eating disorders. *Eating Behaviors*, *14*(2), 233 236. https://doi.org/10.1016/j.eatbeh.2012.12.003

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33 65. https://doi.org/10.1016/j.acn.2004.02.005

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271 276.

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry*, *25*(4), 405 414. https://doi.org/10.2307/2785779

Specht, K., Lie, C. H., Shah, N. J., & Fink, G. R. (2009). Disentangling the prefrontal network for rule selection by means of a non verbal variant of the Wisconsin Card Sorting Test. *Human Brain Mapping*, *30*(5), 1734 1743. https://doi.org/10.1002/hbm.20637

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.

Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., … Morris, R. (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PLoS ONE*, *7*(1). https://doi.org/10.1371/journal.pone.0028331

Vall, E., & Wade, T. D. (2015). Trail Making Task performance in inpatients with anorexia nervosa and bulimia nervosa. *European Eating Disorders Review*, *23*(4), 304 311. https://doi.org/10.1002/erv.2364

Van Autreve, S., De Baene, W., Baeken, C., van Heeringen, C., & Vervaet, M. (2013). Do restrictive and bingeing/purging subtypes of anorexia nervosa differ on central coherence and set shifting? *European Eating Disorders Review*, *21*(4), 308 314. https://doi.org/10.1002/erv.2233

Van Eylen, L., Boets, B., Steyaert, J., Evers, K., Wagemans, J., & Noens, I. (2011). Cognitive flexibility in autism spectrum disorder: Explaining the inconsistencies? *Research in Autism Spectrum Disorders*, *5*(4), 1390 1401. https://doi.org/10.1016/j.rasd.2011.01.025

Wollenhaupt, C., Wilke, L., Erim, Y., Rauh, M., Steins Loeber, S., & Paslakis, G. (2019). The association of leptin secretion with cognitive performance in patients with eating disorders. *Psychiatry Research*, *276*, 269 277. https://doi.org/10.1016/j.psychres.2019.05.001

Wu, M., Brockmeyer, T., Hartmann, M , Skunde, M., Herzog, W., & Friederich, H. C. (2014). Set shifting ability across the spectrum of eating disorders and in overweight and obesity: A systematic review and meta analysis. *Psychological Medicine*, *44*(16), 3365 3385. https://doi.org/10.1017/S0033291714000294

Yuan, P., & Raz, N. (2014). Prefrontal cortex and executive functions in healthy adults: A meta analysis of structural neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 180 192. https://doi.org/10.1016/j.neubiorev.2014.02.005

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2018). Cognitive underpinnings of nationalistic ideology in the context of Brexit. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), E4532 E4540. https://doi.org/10.1073/pnas.1708960115

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

CHAPTER THREE

Chapter 3 is published as:

Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, *88*(102061). https://doi.org/10.1016/j.cpr.2021.102061

This article was reproduced with permission from Elsevier. A signed statement of contribution for this co-authored manuscript can be found in Appendix F.

Clinical Psychology Review 88 (2021) 102061

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Same room - different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults

Caitlin A. Howlett [a], Michael A. Wewege [b,c], Carolyn Berryman [d], Annika Oldach [e], Elizabeth Jennings [a], Emily Moore [a], Emma L. Karran [a], Kimberley Szeto [a], Leander Pronk [a], Stephanie Miles [f], G. Lorimer Moseley [a,*]

[a] Innovation, Implementation & Clinical Translation (IIMPACT) in Health, University of South Australia, Adelaide, Australia
[b] School of Health Sciences, Faculty of Medicine and Health, University of New South Wales, Sydney, Australia
[c] Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, Australia
[d] Brain Stimulation, Imaging and Cognition Research Group, School of Medicine, The University of Adelaide, Adelaide, Australia
[e] School of Human Sciences, Osnabrück University, Osnabrück, Germany
[f] Centre for Mental Health, Swinburne University of Technology, Melbourne, Australia

ARTICLE INFO

Keywords:
Cognitive flexibility
Self-report test
Neuropsychological test
Systematic review
Healthy adults
Correlation

ABSTRACT

Cognitive flexibility can be thought of as the ability to effectively adapt one's cognitive and behavioural strategies in response to changing task or environmental demands. To substantiate the common inference that self-report and neuropsychological tests of cognitive flexibility provide 'different windows into the same room', we undertook a systematic review and meta-analysis to determine whether self-report and neuropsychological tests of cognitive flexibility are related in healthy adults. Ten databases and relevant grey literature were searched from inception. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were adhered to. Twenty-one articles satisfied our inclusion criteria. A multi-level random-effects meta-analysis revealed no relationship (0.05, 95% CI = 0.00 to 0.10). Random-effects meta-analyses raised the possibility that the Cognitive Flexibility Scale and the Trail Making Test – part B (time) may be related (0.19, 95% CI = 0.06 to 0.31). We conclude that the relationship between self-report and neuropsychological tests of cognitive flexibility is not large enough to be considered convincing evidence for the two assessment approaches sharing construct validity. These results have clear implications for assessing and interpreting cognitive flexibility research and clinical practice.

Cognitive flexibility has been described as the ability to effectively adapt one's cognitive and behavioural strategies in response to changing task or environmental demands (Deák, 2003; Diamond, 2013). Cognitive flexibility is thought to emerge from the coherent interaction between higher-order cognitive processes including attention, working memory, inhibition, and switching; all of which are also core processes for executive function (Dajani & Uddin, 2015). Individuals possess different degrees of cognitive flexibility (i.e., higher versus lower levels), with more rigid cognitions and perseverative behaviours often referred to as cognitive *in*flexibility or 'rigidity' (Diamond, 2013; Ionescu, 2012). Deficits in neuropsychological tests, and characteristic patterns of self-reported questionnaire responses, have been observed in people with persistent pain (Berryman et al., 2014; Bunk, Preis, Zuidema, Lautenbacher, & Kunz, 2019) and linked to psychopathology or associated symptoms, including depression (Lee, Hermens, Porter, & Redoblado-Hodge, 2012), anxiety (Lee & Orsillo, 2014), schizophrenia, (Thoma, Wiebel, & Daum, 2007) and eating disorders (Abbate-Daga et al., 2011; Lounes, Khan, & Tchanturia, 2011; Miles, Gnatt, Phillipou, & Nedeljkovic, 2020; Roberts, Tchanturia, Stahl, Southgate, & Treasure, 2007; Steinglass, Walsh, & Stern, 2006; Tchanturia et al., 2012; Tchanturia, Harrison, Davies, Roberts, & Oldershaw, 2011; Westwood, Stahl, Mandy, & Tchanturia, 2016). Higher cognitive flexibility may be protective against severe post-traumatic stress disorder symptoms shortly after trauma (Ben-Zion et al., 2018) – placing cognitive flexibility as a

* Corresponding author at: IIMPACT in Health, University of South Australia, GPO Box 2471, Adelaide, South Australia 5001, Australia.
E-mail address: lorimer.moseley@unisa.edu.au (G.L. Moseley).

https://doi.org/10.1016/j.cpr.2021.102061
Received 13 November 2020; Received in revised form 1 June 2021; Accepted 17 June 2021
Available online 26 June 2021
0272-7358/© 2021 Published by Elsevier Ltd.

potentially important component of resilience under threat conditions.

Definitions and descriptions of cognitive flexibility vary, and there is no unified conceptualisation or accepted operational definition (Dajani & Uddin, 2015; Ionescu, 2012). The term "cognitive flexibility" has been used synonymously with *mental flexibility*, *psychological flexibility* and *set-shifting* (e.g., Diamond, 2013; Palm & Follette, 2011; Wecker, Kramer, Hallam, & Delis, 2005); set-shifting has also been proposed as a sub-component of cognitive flexibility (Lounes et al., 2011). Ionescu (2012) identified four definitions of cognitive flexibility that are in common use in the literature: (i) a specific ability or skill like "set-shifting", (ii) an ability of higher-order executive function, (iii) a property of cognitive states, and (iv) an outcome of divergent thinking. Both psychological flexibility – a concept derived from Acceptance and Commitment Therapy (ACT; Hayes, Luoma, Bond, Masuda, & Lillis, 2006) and specifically built around the objectives to be open, be present and do what matters in the context of self (Hoffmann, Rask, & Frostholm, 2019) – and cognitive flexibility, have both been previously co-associated with psychopathology (Berman, Wheaton, McGrath, & Abramowitz, 2010; Tchanturia et al., 2004; Westwood et al., 2016). It has been proposed that psychological and cognitive flexibility share conceptual overlap, and perhaps underlying neuronal circuitry (Whiting, Deane, Simpson, McLeod, & Ciarrochi, 2017); however, while Whiting et al. (2017) conceptualised cognitive flexibility as a subcomponent, but not a pre-requisite of psychological flexibility, others have used psychological and cognitive flexibility interchangeably (Palm & Follette, 2011).

The variable definitions, understanding, and usage of cognitive flexibility extends to its assessment. There are two broad approaches — self-report and neuropsychological testing. The self-report tests most commonly used to assess cognitive flexibility appear to be the Cognitive Flexibility Scale (CFS; Martin & Anderson, 1998; Martin & Rubin, 1995), the Cognitive Flexibility Inventory (CFI; Dennis & Vander Wal, 2010) and the Behaviour Rating Inventory of Executive Function-Adult version (BRIEF-A (shift subscale); Roth, Isquith, & Gioia, 2005). Self-report tests of cognitive flexibility are favoured clinically because they are relatively easy to administer, and are time and cost effective (Dennis & Vander Wal, 2010; Johnco, Wuthrich, & Rapee, 2014). However, like other self-report tests, they are vulnerable to reporting bias, rely on an individual's perception of their abilities, and will inevitably involve other processes of executive function.

Neuropsychological tests overcome some of the limitations of self-report tests including social desirability bias and the capacity for self-awareness. The most widely recognised neuropsychological tests of cognitive flexibility are the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948; Heaton, Chelune, Talley, Kay, & Curtiss, 1993), the Trail Making Test (TMT; Reitan, 1958), and the Stroop Colour-Word Test (Stroop; Golden, 1978; Stroop, 1935) – ranked in the top 40 most commonly administered neuropsychological assessments by clinical neuropsychologists (Rabin, Barr, & Burton, 2005). Neuropsychological tests that are used to assess cognitive flexibility are also used to more broadly assess executive function (e.g., working memory and inhibition), further highlighting the variability and ambiguity in the field (Miyake, Emerson, & Friedman, 2000). Despite their popularity, neuropsychological tests may not fully capture aspects of cognitive flexibility that allow targeting of psychotherapy interventions (Dennis & Vander Wal, 2010). Further, neuropsychological tests such as the WCST are challenging to administer and score, especially if using non-computerised versions of the task (Miles et al., 2021), and have shown to be susceptible to practice effects (Basso, Bornstein, & Lang, 1999; Basso, Lowery, Ghormley, & Bornstein, 2001).

The variability and lack of consensus surrounding cognitive flexibility may underpin the common assumption that self-report and neuropsychological assessments of cognitive flexibility provide 'different windows into the same room'. That is, they are often conflated, used interchangeably, and are thought to interrogate the same broad construct (Friederich & Herzog, 2011; Fujino et al., 2019; Holliday, Tchanturia, Landau, Collier, & Treasure, 2005; Perpiná, Segura, & Sánchez-Reales, 2017; Tchanturia et al., 2012). This assumption is important because it underpins the use of self-report assessments in clinical practice and the extrapolation of neuropsychological tests to broader clinical notions of cognitive and psychological flexibility. The importance of this assumption is further exemplified when faced with assessing the effects of interventions designed to improve cognitive flexibility (Baker et al., 2018; Ben-Zion et al., 2018; Leppanen, Adamson, & Tchanturia, 2018). Preliminary findings from three studies produced contrasting results when evaluating the relationship between several pairs of tests designed to assess cognitive flexibility (Gonzalez, Figueroa, Bellows, Rhodes, & Youmans, 2013; Johnco et al., 2014; Lounes et al., 2011); however, the strength of the relationship between self-report and neuropsychological tests of cognitive flexibility across the entire evidence base of healthy adults remains unknown.

The aim of the current systematic review was to determine whether, and to what extent, self-report and neuropsychological tests of cognitive flexibility are related in healthy adults. By doing so, we aimed to substantiate the implicit assumption that the two assessment approaches interrogate the same broad construct. Such substantiation would support the use of self-report questionnaires as providing valid insight into cognitive processing challenges, and the extrapolation of neuropsychological tests to clinical assessment and psychotherapy in people who report cognitive and behavioural rigidity.

1. Method

1.1. Data sources and search strategy

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement guidelines (Moher et al., 2009; Shamseer et al., 2015; see Fig. 1). Prior to commencing our database searches, we searched within the International Prospective Register of Systematic Reviews (PROSPERO) database for possible registered protocols that may have been conducting the same systematic review. No such protocol was identified on the database, thus a protocol was prospectively registered on PROSPERO (registration number: CRD42019131567; available from https://www.crd.york.ac.uk/prospero/display_record.php?ID CRD42019131567), and later updated and registered on Open Science Framework (OSF) (https://osf.io/7yzmq/). Any deviations from the registered protocol were logged and are available on OSF. One reviewer (CAH) executed the search strategy from inception to 23rd October 2020. Searches were performed in the following databases: MEDLINE (via OvidSP), EMBASE (via OvidSP), EMCARE (via OvidSP), SCOPUS (via Elsevier), PsycINFO (via OvidSP), Web of Science Core Collection (via Clarivate Analytics), Cochrane Central Register of Controlled Trials (CENTRAL), the Cochrane Database of Systematic Reviews, PEDro, and OT seeker. The general search strings were (cognitive* flexib* OR mental* flexib* OR psychological* flexib* OR wisconsin card? sort* OR trail mak*) AND (cognitive flexibility scale? OR cognitive flexibility inventor* OR self?report*), and Medical Subject Headings (MeSH) were used where possible. The search strategy was devised in accordance with the PICO framework (Schardt, Adams, Owens, Keitz, & Fontelo, 2007) and the final strategy was finalised in collaboration with an experienced academic librarian (see Appendix A for the full electronic search strategy in Medline and Supplementary File S1 for search strategies used for all other databases and grey literature). Search strings were modified to accommodate for the functionality of each database.

Grey literature was searched by one reviewer (CAH) from inception to 29th August 2019, and comprised of dissertation databases (Trove, ProQuest), clinical trial registries (Australian New Zealand Clinical Trials Registry (ANZCTR), Clinicaltrials.gov, World Health Organization International Clinical Trials Registry Platform (WHO ICTRP)), Google Scholar (first 100 references) and conference proceedings. All search

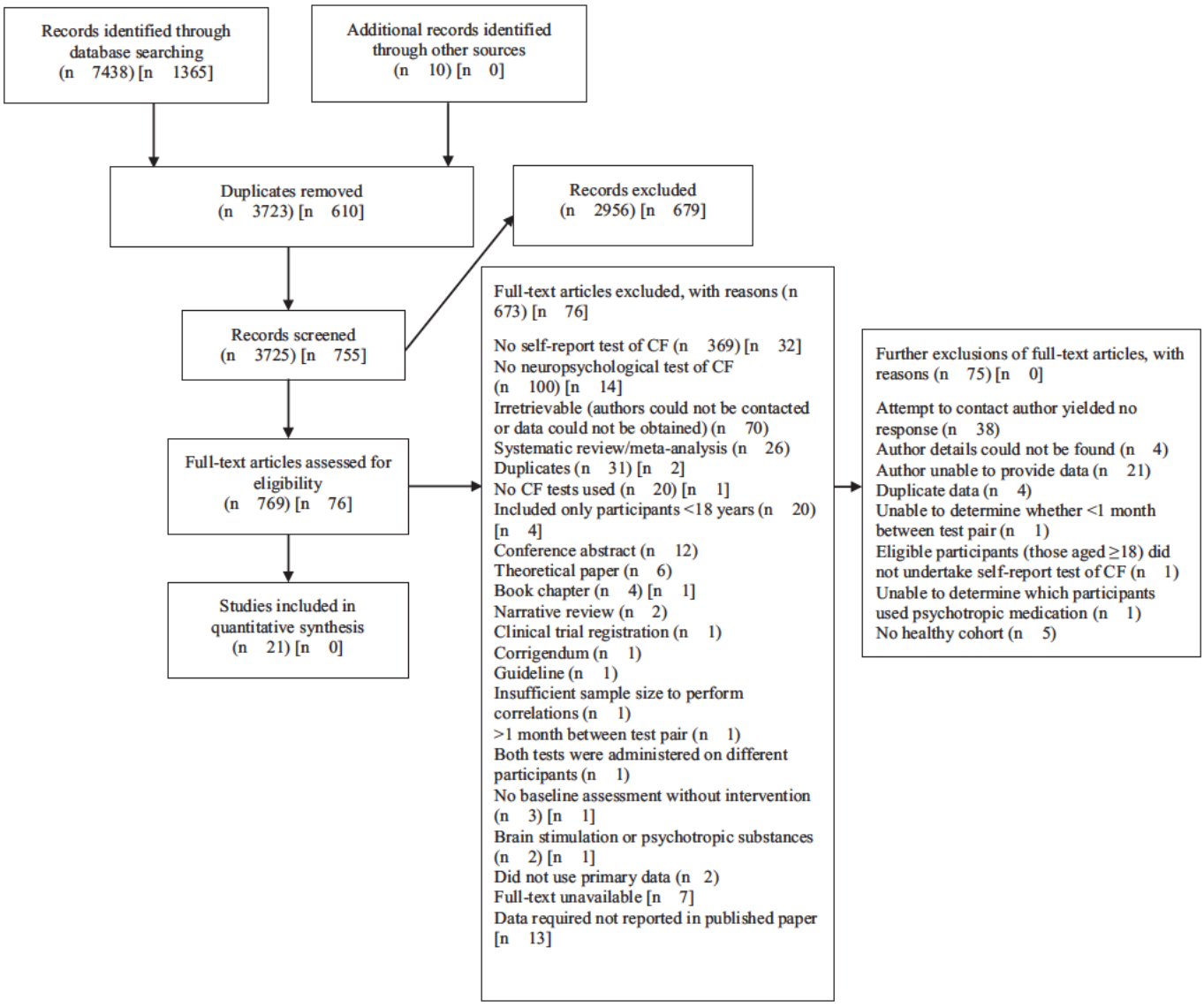


Fig. 1. PRISMA flow diagram.
Note. (n =) original search, [n =] updated search.

results were limited to studies with human participants but there were no restrictions on language, country of origin, date of publication nor publication status. Key authors in the field were identified via SCOPUS and contacted to ensure that no relevant studies were missed in the search strategy. The reference lists of relevant systematic reviews and of all included studies were manually searched to identify any additional records. For articles with titles that appeared relevant, full-text records were retrieved and screened for eligibility.

1.2. Study eligibility

To be eligible for inclusion in this review, all studies had to either evaluate the relationship between, or collect the data for, at least one self-report test of cognitive flexibility, for example – the Cognitive Flexibility Inventory – and at least one neuropsychological test of cognitive flexibility, for example – the Wisconsin Card Sorting Test. Articles were not excluded based on study design, providing the design permitted correlations to be generated (i.e., $n > 1$). Studies that included additional tests to which we were naïve but explicitly stated that the test was an assessment of flexibility (*a priori*) in the introduction, and/or methods section of the paper were included. Those studies in which flexibility was used as an interpretation of results *post-hoc* were not included in the review. For tests that were recognised to be assessing cognitive flexibility, we decided *a priori* which outcome variables to select on the basis of previous literature. For example, for the WCST, we decided to prioritise the outcome variables perseverative errors and perseverative responses since these variables are the most widely accepted outcomes of cognitive flexibility within the field, and there is a lack of consensus as to whether other outcomes of the WCST (e.g., failure-to-maintain-set) assess cognitive flexibility (Figueroa & Youmans, 2013; Miles et al., 2021). For tests to which we were naïve to, we discussed the best outcome(s) on the basis of the article that used it and we sought clarification with the author(s) where necessary.

Studies were ineligible for inclusion if they (i) included participants <18 years of age; (ii) did not use at least one self-report test and at least one neuropsychological test of cognitive flexibility; (iii) included clinical cohorts (iv) were unable to provide correlational or raw data where it was not reported in the original source; (v) had participants undertake the two forms of tests more than one month apart, or separated by an intervention; (vi) administered the two tests on

different participants (i.e., assumed equivalence between samples); or (vii) used psychotropic substances or neural stimulation prior to administering the tests. Commentaries, conference abstracts, editorials, protocols, and letters to the editor were only included if a full-text article could be sourced and/or if full data were able to be obtained from the author(s). Book chapters and secondary data (i.e., systematic reviews or meta-analyses) were excluded from this review.

1.3. Article selection

All records identified by the search strategy were exported into Endnote (Clarivate Analytics, Philadelphia) and the records were then imported into Covidence (Veritas Health Innovation Ltd., Australia), an online screening platform, where duplicates were removed. Six independent reviewers (CAH, LP, EJ, SM, EM, KS) screened the records by title and abstract in pairs to determine article eligibility against the predetermined criteria. Irrelevant citations were removed. Articles potentially meeting the criteria were progressed to full-text screening, and those that lacked sufficient information, and thus eligibility could not be determined were retrieved and formally assessed during full-text review. Multiple reviewers independently screened full-text articles in pairs (CAH, LP, EJ, SM, AO, EM, and KS). Disagreements were either resolved by a third reviewer (CAH, EM or KS) or through discussion (with GLM) if consensus could not be adequately reached. In instances where entire conference proceedings were retrieved, key terms including "flexib", "wisconsin", "trail mak" and "stroop" were searched within the proceedings and authors were contacted to request full-texts of abstracts that appeared relevant or where there was insufficient information to determine inclusion.

Additionally, where full-texts were unavailable, authors were contacted to clarify eligibility and/or to provide full-texts. A total of 108 authors were contacted by either email or ResearchGate (ResearchGate, Berlin, Germany) and four authors were contacted via phone. All authors were given two weeks to provide either published papers, unpublished manuscripts and/or the raw or aggregate data. Of 112 authors contacted, 70 records were irretrievable because either the author(s) could not be found, attempts to contact did not return any response, or they were unable to provide the necessary data. The remaining 39 authors replied confirming ineligibility and three authors confirmed eligibility; however, only one author was able to provide sufficient data and thus could be included (Clayson, 2018). Any non-English publications were translated using Google Translate, but none were eligible for inclusion.

1.4. Risk of bias assessment

Study quality was assessed independently by two reviewers (CAH [all papers] and LP, AO, EM, or KS), using a customised risk of bias form. The form was developed *a priori*, adapted from relevant items from the Appraisal tool for Cross-Sectional Studies (Downes, Brennan, Williams, & Dean, 2016) and the National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-sectional Studies (National Heart, Lung, and Blood Insitute, 2014), and piloted prior to data collection. Two reviewers completed this form independently for each study and then assessments were compared. Any disagreements were resolved by discussion between the two reviewers, and a third reviewer (GLM) was consulted when necessary.

1.5. Data extraction

Two independent reviewers (CAH [all papers] and either LP, AO, EM or KS) used a piloted, customised data extraction form to extract the relevant data, and the results were compared for accuracy. The following data were extracted from included records: citation details, study characteristics (aim(s), design, country, language in which tests were administered, study conditions, setting, recruited sample size and sample size analysed, participant inclusion/exclusion criteria, participant characteristics (sex and/or gender, health status, means and standard deviations for age and years of education), assessment pair details [name of self-report test (e.g., Cognitive Flexibility Scale), name of neuropsychological test (e.g., Wisconsin Card Sorting Test), outcome measure of test (e.g., number of perseverative errors), direction of test outcome (e.g., higher scores indicate better performance)], length of time between administration of each test in the pair, assessment time point (i.e., no intervention, pre-intervention, post-intervention), intervention details (if applicable), authors' definition of cognitive flexibility (if reported), results of correlation(s) (e.g., Pearson's *r* or Spearman's rho, *p*-values, confidence intervals and/or standard errors), descriptives of assessments (mean and standard deviation), and the reporting of power analyses. Due to time constraints, authors of dissertations were contacted prior to performing data extraction and were given one week to respond in order to be included in the review. Once the data of all studies were extracted, if information beyond that contained in the full-text article was required, three attempts were made to contact the authors over six weeks. If all three attempts were unsuccessful or the authors were unable to provide sufficient information for analyses, the missing information was considered irretrievable and the study was not included in this review.

Where the raw data were provided by authors, correlations were performed and analysed using IBM SPSS (version 20; IBM Corp, 2011); this process was undertaken independently by two reviewers (CAH and KS). Where raw data were obtained from authors, statistical assumptions were assessed following the procedures outlined by Kim (2013). In instances where assumptions of normality were violated, Spearman's rho correlation coefficients were used instead of Pearson's *r* coefficients.

1.6. Data analysis and data pooling

We performed a multi-level random-effects meta-analysis using R (R Core Team, 2019) and the *metafor* package (Viechtbauer, 2010) (https://osf.io/6qs2u/). This was a necessary deviation from our *a priori* protocol because RevMan 5.3 software (The Cochrane Collaboration, 2014) did not allow the required analyses, but a multi-level meta-analysis enables the analysis of all eligible effects from included studies, accounting for correlations between outcomes performed by the same participants. In this analytical approach, Level 1 variance represented within-study sampling variance, Level 2 represented variance between effect sizes within a study, and Level 3 represented variance between studies. We fitted the model via restricted maximum-likelihood estimation with a compound symmetry variance structure and Knapp and Hartung's adjustment (Assink & Wibbelink, 2016). Correlation coefficients were transformed to Fisher's z correlation coefficients for analysis with the *esc* package (Lüdecke, 2019) and converted back to correlation coefficients for presentation in forest plots (Borenstein, Hedges, Higgins, & Rothstein, 2009). Where the scores for the test pair were going in opposite directions (i.e., higher scores equal worse performance on one test and higher scores equal better performance on the other test), we multiplied one of the test outcomes by −1. This reversed the scale of the scores such that both outcomes for the test pair faced the same direction (i.e., both negative or both positive).

We also performed a random-effects meta-analysis for a test pair when at least two studies used the same outcomes of a particular pair (e.g., BRIEF-A (shift subscale) and WCST (perseverative errors)), or when one study used the same test pair with two different groups (i.e., comparing males and females). Effect sizes were interpreted using the conventions outlined by Cohen (1992). Effects of 0.1, 0.3 and >0.5 were considered small, moderate and large, respectively. Statistical significance was set at $p < .05$. Heterogeneity was considered statistically significant when the χ^2 statistic (otherwise known as Q) had $p < .10$, and substantial when the I^2 statistic was >50% (Higgins et al., 2019). To identify potential differences in effect estimates between studies, a moderator analysis was performed by age (above and below

60 years), and an *a priori* sensitivity analysis was conducted to consider studies without a conflict of interest statement.

1.7. Moderator and sensitivity analyses

We planned a subgroup analysis on age because there is some evidence that cognitive flexibility may change over the life span. Ashendorf and Mccaffrey (2008) found that older adults score lower on a range of outcome variables of the WCST than younger adults, and that older age was positively associated with scores on WCST variables including perseverative errors and perseverative responses. Similarly, Wecker et al. (2005) demonstrated that performance on verbal and non-verbal tests of the Delis-Kaplan Eexecutive Function System (D-KEFS) battery were influenced by age even after controlling for demographic factors and component skills. Thus, both studies concluded that declines in cognitive flexibility occur with increasing age. We planned to perform a subgroup analysis on participants aged above and below 60 years, and we chose this cut-off based on previous literature that suggests that declines in cognitive flexibility may be observed (Gajewski, Ferdinand, Kray, & Falkenstein, 2018; Nguyen, Murphy, & Andrews, 2019). On the basis of peer review, we modified this plan to instead undertake a moderator analysis on age. We also planned a sensitivity analysis removing studies that included a statement declaring conflicts of interest because failing to do so is associated with an established risk of reporting false positive findings (Resnik, 2007).

2. Results

2.1. Retrieved studies

Our search strategy identified 8813 articles. After removing duplicates, 4480 articles were screened by title and abstract according to our inclusion/exclusion criteria. Eight hundred and forty-five full-text articles were retrieved and screened for eligibility. Of these, 749 were excluded (see Fig. 1 for a summary of reasons for exclusion). After completing full-text screening, a total of 96 articles required author correspondence to confirm eligibility. All 96 articles at this stage required either 1) further clarification regarding the eligibility criteria and/or 2) requests for data (aggregate or raw). Of all 96 articles for whose authors we contacted, we excluded an additional 75 articles for various reasons (see Fig. 1), leaving 21 articles eligible for inclusion. Full details on attempts to obtain data through author correspondence are provided in the Supplementary Files (see Supplementary File S2).

A total of 21 articles were included in this review after (i) two independent reviewers agreed on article eligibility or article authors were able to confirm eligibility, and (ii) aggregate or raw data were either in the article or could be obtained from authors. Raw data were obtained from the authors of seven articles (Alfimova, Golimbet, Lezheiko, & Korovaitseva, 2017; Bellinger et al., 2015; Dassen, Houben, Allom, & Jansen, 2018; Dawson, Shear, & Strakowski, 2012; Meltzer et al., 2017; Reed, Combs, & Segerstrom, 2020; Zmigrod, Rentfrow, & Robbins, 2018) and aggregate data were obtained from the authors of six articles (Clayson, 2018; Cohen, Jensen, Stange, Neuburger, & Heimberg, 2017; Dingemans, Vanhaelen, Aardoom, & van Furth, 2019; Johnco et al., 2014; Kuula et al., 2018; McCloskey, 2010). The remaining eight articles (Aytug, Rua, Brazeal, Almaraz, & González, 2018; Brenner, 2017; Drake, 2008; Hertzberg, 2017 (Sample A and B); Lau, 2012; Lounes et al., 2011; Saylik & Szameitat, 2018) contained sufficient information to extract the required data without contacting authors; however, any of the eligibility criteria that were unclear from the article were confirmed with the authors. Where studies included (i) participants who were under 18 years of age or (ii) participants who did not complete the test pair(s) within one month, we requested that these participants be removed from the dataset where aggregate data was provided. Where the age range of the sample was not reported in the original article, we contacted the authors and requested that the provided aggregate data only include participants who were ≥18 years; where raw data were obtained, we omitted participants <18 years of age from the dataset prior to analyses. These changes mean that the data presented in this review can vary from the data presented in the original articles/dissertations.

2.2. Risk of bias of included studies

All studies were graded as having moderate to high risk of bias (see Table 4), primarily due to the poor quality of the methods and procedures, and insufficient reporting. The aims and/or research questions of four studies were aligned with the aims of this review (Brenner, 2017; Hertzberg, 2017 (Sample B); Johnco et al., 2014; Lounes et al., 2011). We conducted *post-hoc* power calculations using the online calculator by Lachin (1981) and aimed to detect medium effect sizes with α 0.05 and β 0.20 (power 0.80) (Cohen, 1992). With the exception of Aytug et al. (2018), the *post-hoc* power calculations revealed that none of the studies were adequately powered to detect the adjusted power estimates. No studies reported registering a protocol *a priori*; eight studies declared funding sources; ten studies declared conflicts of interests. All studies, with the exception of Drake (2008), either inadequately reported or did not counterbalance the order of test administration across participants. Only two studies (Brenner, 2017; McCloskey, 2010) reported that all of the cognitive flexibility tests used were reliable and provided reliability coefficients. All studies, except Aytug et al. (2018) and Hertzberg (2017) (Sample A), either adequately described the results of the correlations or the information was obtained through correspondence with the authors. For all included studies, sample demographics were adequately described or obtained. The majority of the studies were internally consistent (i.e., free of inconsistencies within the paper), and descriptive data of test outcomes were adequately described or able to be obtained from authors for most of the studies. Only two studies reported planned correlational analyses in their methods *a priori* and most studies adequately reported that the assumptions of normality were met. The findings were mixed for the study conditions being consistent across participants.

2.3. Study characteristics

Of the 21 included studies (n 2762), eight used a cross-sectional design, two used a longitudinal design, nine combined correlational and cross-sectional methods, one was a retrospective medical chart review, and one was a correlational, cross-sectional pilot study. Cross-sectional studies that sought to understand associations between variables were considered *cross-sectional and correlational;* cross-sectional studies that did not seek to investigate associations between variables were considered merely *cross-sectional.* Sample sizes ranged from 8 to 687; the largest sample for a single correlation was n 386. There were 1617 females, 1142 males and one person who identified as transgender across all cohorts. The sample weighted mean age of the included studies was 34.9 (SD 6.4), ranging from 19 to 81 years. Table 1 summarises the key characteristics of the included studies.

2.4. Summary of self-report and neuropsychological assessments of cognitive flexibility

2.4.1. Self-report assessments of cognitive flexibility

Eight different self-report tests of cognitive flexibility were used across all included studies (see Fig. 2). The shift subscale of the BRIEF-A (Guy, Isquith, & Gioia, 2004) was the most commonly used self-report test of cognitive flexibility, with 11 studies (52%; Alfimova et al., 2017; Bellinger et al., 2015; Clayson, 2018; Dassen et al., 2018; Dawson et al., 2012; Dingemans et al., 2019; Hertzberg, 2017 (Sample A and B); Kuula et al., 2018; Meltzer et al., 2017; Reed et al., 2020) using this test. Five studies (Aytug et al., 2018; Brenner, 2017; Johnco et al., 2014; Lau, 2012; Lounes et al., 2011) used the CFS (Martin & Anderson, 1998;

Table 1
Study characteristics of included studies.

Citation & year	Country of origin	Study design	N	Mean age (SD)	Gender or sex	Mean years of education (SD)
Alfimova et al. (2017)	Russia	Cross-sectional, correlational	69	40.6 (13.3)	21 M; 48 F	15.3 (2.7)
Aytug et al. (2018)	United States of America	Cross-sectional	242	23.2 (5.2)	112 M; 131 F	–
Bellinger et al. (2015)	United States of America	Cross-sectional	8	18.7 (0.8)	1 M; 7 F	–
Brenner (2017)	United States of America	Cross-sectional, correlational	88	36.7 (12.3)	88 M	–
Clayson (2018)	United States of America	Cross-sectional	38	20.5 (3.5)	38 F	–
Cohen et al. (2017)	United States of America	Longitudinal	142	47.6 (15.9)	44 M; 98 F	–
Dassen et al. (2018)	The Netherlands	Cross-sectional, correlational	71	43.4 (13.4)	16 M; 55 F	–
Dawson et al. (2012)	United States of America	Cross-sectional, correlational	85	19.5 (2.9)	36 M; 49 F	–
Dingemans et al. (2019)	The Netherlands	Cross-sectional	55	36.5 (12.3)	6 M; 49 F	–
Drake (2008)	United States of America	Cross-sectional	172	18.9 (2.3)	55 M; 117 F	–
Hertzberg (2017) (Sample A)	United States of America	Pilot study	15	24.5 (8.4)	5 M; 10 F	–
Hertzberg (2017) (Sample B)	United States of America	Retrospective medical chart review	98	26.0 (11.6)	46 M; 52 F	–
Johnco et al. (2014)	Australia	Cross-sectional, correlational	49	67.6 (6.3)	13 M; 36 F	14.4 (3.1)
Kuula et al. (2018)	Finland	Cross-sectional	687	25.4 (0.6)	301 M; 386 F	–
Lau (2012)	United States of America	Cross-sectional	92	33.3 (9.9)	35 M; 56 F; 1 T	–
Lounes et al. (2011)	The United Kingdom	Cross-sectional, correlational	49	24.1 (4.4)	49 F	–
McCloskey (2010)	United States of America	Cross-sectional, correlational	75	20.9 (2.3)	36 M; 39 F	–
Meltzer et al. (2017)	United States of America	Cross-sectional	194	80.9 (5.3)	65 M; 129 F	15.1 (3.2)
Reed et al. (2020)	United States of America	Longitudinal	148	79.5 (5.4)	63 M; 85 F	16.4 (2.5)
Saylik & Szameitat (2018)	The United Kingdom	Cross-sectional, correlational	90	22.5 (6.2)	47 M; 44 F	–
Zmigrod et al. (2018)	The United Kingdom	Cross-sectional, correlational	295	37.8 (13.5)	152 M; 139 F	–

Note. F = female; M = male; N = sample size; SD = standard deviation; T = a person who identifies as transgender; – = information not reported or could not be obtained.

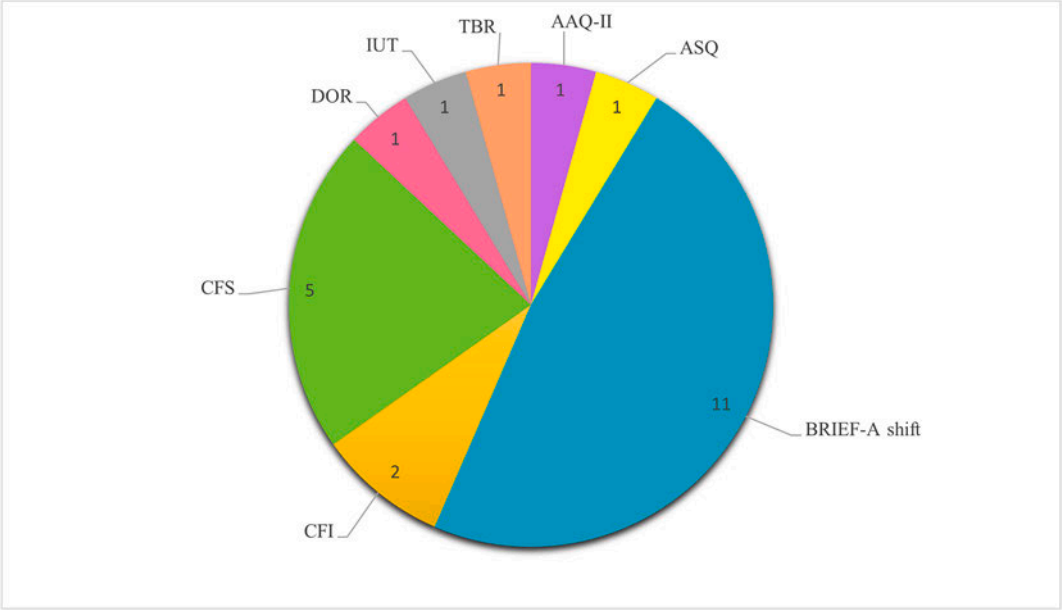


Fig. 2. Number of included studies that used each self-report test of cognitive flexibility.
Note. AAQ-II = Acceptance and Action Questionnaire-II; ASQ = Attributional Style Questionnaire; BRIEF-A shift = Behaviour Rating Inventory of Executive Function-Adult shift; CFI = Cognitive Flexibility Inventory; CFS = Cognitive Flexibility Scale; DOR = Dependence on Routines scale; IUT = Intolerance for Uncertainty scale; TBR = Test of Behavioural Rigidity.

Martin & Rubin, 1995), whilst two studies (Cohen et al., 2017; Johnco et al., 2014) used the CFI (Dennis & Vander Wal, 2010). One study (Zmigrod et al., 2018) used both the Intolerance for Uncertainty scale (Carleton, Norton, & Asmundson, 2007) and the Dependence on Routines scale. One study (Saylik & Szameitat, 2018) used the Attributional Style Questionnaire (Peterson et al., 1982); one study (Drake, 2008) used the Acceptance and Action Questionnaire II (Bond, Hayes, Baer, & A., 2011); and one study (McCloskey, 2010) used the Questionnaire subtest of the Test of Behavioural Rigidity (Schaie & Parham, 1975). Table 2 summarises key assessment details of the included studies.

2.4.2. *Neuropsychological assessments of cognitive flexibility*

Thirteen different neuropsychological tests of cognitive flexibility were used across the included studies (see Fig. 3). Of these, the TMT (Reitan, 1958) was the most commonly used neuropsychological test of cognitive flexibility, with 11 studies (52%; Alfimova et al., 2017; Brenner, 2017; Dassen et al., 2018; Dingemans et al., 2019; Hertzberg, 2017 (Sample A and B); Johnco et al., 2014; Kuula et al., 2018; Lau, 2012; Meltzer et al., 2017; Reed et al., 2020) using this test. Six studies (Clayson, 2018; Dingemans et al., 2019; Hertzberg, 2017 (Sample B); Johnco et al., 2014; Lau, 2012; Zmigrod et al., 2018) used the WCST (Grant & Berg, 1948; Heaton et al., 1993) and six studies (Alfimova et al., 2017; Clayson, 2018; Dingemans et al., 2019; Johnco et al., 2014; Kuula et al., 2018; Meltzer et al., 2017) used the Stroop test (Golden, 1978; Stroop, 1935). Three studies (Johnco et al., 2014; Meltzer et al., 2017; Reed et al., 2020) used the Controlled Oral Word Association Test

Table 2
Assessment details and correlational data of included studies.

Citation & year	Language test(s) were administered in	Self-report test	Neuropsychological test	N in correlation	Raw correlation (*r or rho*)	95% CI or SE	*p*-value
[1]Alfimova et al. (2017)	BRIEF-A, Stroop, TMT & Verbal Fluency; Russian versions	BRIEF-A/shift subscale	TMT/part B time	69	*rho* 0.248	95% CI: 0.033, 0.461	*p* .040
		BRIEF-A/shift subscale	Stroop/Colour-Word Trial*	69	*r* 0.081	95% CI: −0.113, 0.279	*p* .508
		BRIEF-A/shift subscale	Stroop/Golden Interference Score	69	*r* 0.216	95% CI: 0.041, 0.383	*p* .075
Aytug et al. (2018)		[2]CFS/total score	[2]AUT/flexibility	242	0.29 (*)	95% CI: 0.171, 0.401	*p* < .01
[1]Bellinger et al. (2015)		BRIEF-A/shift subscale*	D-KEFS Verbal Fluency/switching accuracy*	8	*r* 0.243	95% CI: −0.447, 0.772	*p* .562
		BRIEF-A/shift subscale*	D-KEFS Verbal Fluency/switching vs. category*	8	*r* 0.435	95% CI: −0.430, 0.960	*p* .282
Brenner (2017)		CFS/total score	TMT/part B time	88	*r* −0.255	95% CI: −0.44, −0.049	*p* .009
[1]Clayson (2018)		BRIEF-A/shift subscale*	WCST/% perseverative errors	38	*r* 0.10	95% CI: −0.23, 0.41	*p* > .05
		BRIEF-A/shift subscale*	WCST/% perseverative responses	38	*r* 0.16	95% CI: −0.17, 0.46	*p* > .05
		BRIEF-A/shift subscale*	Stroop/Interference Score/ incongruent trials – neutral trials	38	*r* −0.04	95% CI: −0.35, 0.28	*p* > .05
[1]Cohen et al. (2017)		CFI/total score (T1)*	Set-shifting task/shift costs (T1)	148	*r* 0.167	95% CI: −0.038, 0.364	*p* .042
		CFI/total score (T2)*	Set-shifting task/shift costs (T2)	116	*r* 0.245	95% CI: 0.078, 0.399	*p* .008
		CFI/total score (T3)*	Set-shifting task/shift costs (T3)	112	*r* −0.018	95% CI: −0.196, 0.144	*p* .853
		CFI/total score (T4)*	Set-shifting task/shift costs (T4)	75	*r* 0.261	95% CI: 0.097, 0.413	*p* .024
		CFI/alternatives subscale (T1)	Set-shifting task/shift costs (T1)	148	*r* 0.104	95% CI: −0.087, 0.299	*p* .21
		CFI/alternatives subscale (T2)	Set-shifting task/shift costs (T2)	116	*r* 0.202	95% CI: 0.024, 0.365	*p* .03
		CFI/alternatives subscale (T3)	Set-shifting task/shift costs (T3)	112	*r* −0.065	95% CI: −0.229, 0.085	*p* .495
		CFI/alternatives subscale (T4)	Set-shifting task/shift costs (T4)	75	*r* 0.207	95% CI: 0.045, 0.345	*p* .074
		CFI/control subscale (T1)	Set-shifting task/shift costs (T1)	148	*r* 0.185	95% CI: −0.024, 0.379	*p* .024
		CFI/control subscale (T2)	Set-shifting task/shift costs (T2)	116	*r* 0.220	95% CI: 0.066, 0.396	*p* .018
		CFI/control subscale (T3)	Set-shifting task/shift costs (T3)	112	*r* 0.052	95% CI: −0.133, 0.215	*p* .587
		CFI/control subscale (T4)	Set-shifting task/shift costs (T4)	75	*r* 0.255	95% CI: 0.04, 0.444	*p* .027
[1]Dassen et al. (2018)	BRIEF-A; Dutch version	BRIEF-A/shift subscale	[2]TMT/part B time*	71	*rho* 0.261	95% CI: 0.010, 0.495	*p* .028
		BRIEF-A/shift subscale	TMT/ part B-A time	71	*rho* 0.206	95% CI: −0.045, 0.449	*p* .085
		BRIEF-A/shift subscale	[2]TMT/part B/A time*	71	*r* 0.125	95% CI: −0.126, 0.367	*p* .299
[1]Dawson et al. (2012)		BRIEF-A/shift subscale*	[2]D-KEFS/inhibition	85	*r* −0.288	95% CI: −0.467, −0.084	*p* .008
		BRIEF-A/shift subscale*	[2]D-KEFS/inhibition/switching	85	*rho* −0.211	95% CI: −0.405, −0.010	*p* .053
[1]Dingemans et al. (2019)	BRIEF-A, Stroop, TMT, WCST; Dutch versions	BRIEF-A/shift subscale*	[2]TMT/part B-A time*	55	*r* −0.247	95% CI: −0.48, 0.019	*p* .069
		BRIEF-A/shift subscale*	TMT/part B/A time	55			

(*continued on next page*)

Table 2 (*continued*)

Citation & year	Language test(s) were administered in	Self-report test	Neuropsychological test	N in correlation	Raw correlation (*r or rho*)	95% CI or SE	*p*-value
					rho −0.245	95% CI: −0.479, 0.021	*p* .072
		BRIEF-A/shift subscale*	[2]TMT/part B time*	55	*rho* −0.327	95% CI: −0.545, −0.068	*p* .015
		BRIEF-A/shift subscale*	[2]WCST/perseverative errors	55	*rho* 0.097	95% CI: −0.172, 0.353	*p* .482
		BRIEF-A/shift subscale*	[2]WCST/perseverative responses*	55	*rho* 0.110	95% CI: −0.159, 0.364	*p* .425
		BRIEF-A/shift subscale*	[2]Stroop/Interference Score/trial 3 – trial 1*	55	*rho* 0.165	95% CI: −0.104, 0.412	*p* .229
		BRIEF-A/shift subscale*	Stroop/trial 3/trial 1	55	*rho* 0.169	95% CI: −0.10, 0.415	*p* .217
Drake (2008)		AAQ-II/total score	[2]IRAP/self-positive blocks	172	*r* 0.048	95% CI: −0.102, 0.196	ns
		AAQ-II/total score	[2]IRAP/self-negative blocks	172	*r* 0.036	95% CI: −0.114, 0.184	ns
Hertzberg (2017) (Sample A)		[2]BRIEF-A/shift subscale	[2]TMT/part B time (standard score)	15	−0.12 (^)	95% CI: −0.595, 0.417	ns
Hertzberg (2017) (Sample B)		BRIEF-A/shift subscale	TMT/part B (standard score)	98	*r* −0.10	95% CI: −0.292, 0.10	ns
		BRIEF-A/shift subscale	[2]WCST/perseverative errors (standard score)	98	*r* −0.11	95% CI: −0.301, 0.09	ns
		BRIEF-A/shift subscale	[2]WCST/perseverative responses (standard score)	98	*r* −0.11	95% CI: −0.301, 0.09	ns
[1]Johnco et al. (2014)		CFI/total score	TMT/part B time	49	*r* −0.223	95% CI: −0.474, 0.041	*p* .123
		CFI/total score	[2]TMT/part B-A time*	49	*r* −0.101	95% CI: −0.377, 0.196	*p* .492
		CFI/total score	[2]TMT/part B/A time*	49	*r* 0.047	95% CI: −0.228, 0.312	*p* .749
		CFI/total score	Stroop/Colour-Word Trial	49	*r* 0.120	95% CI: −0.076, 0.343	*p* .412
		CFI/total score	[2]Stroop/Interference Score*	49	*r* −0.018	95% CI: −0.295, 0.255	*p* .902
		CFI/total score	COWAT/sum total score	49	*r* 0.243	95% CI: −0.006, 0.483	*p* .093
		CFI/total score	WCST/perseverative errors	49	*r* −0.201	95% CI: −0.442, 0.056	*p* .166
		CFI/total score	WCST/% perseverative responses*	49	*r* −0.176	95% CI: −0.434, 0.085	*p* .226
		CFI/alternatives subscale	TMT/part B time	49	*r* −0.056	95% CI: −0.268, 0.146	*p* .704
		CFI/alternatives subscale	[2]TMT/part B-A time*	49	*r* 0.063	95% CI: −0.156, 0.290	*p* .669
		CFI/alternatives subscale	[2]TMT/part B/A time*	49	*r* 0.174	95% CI: −0.034, 0.385	*p* .233
		CFI/alternatives subscale	Stroop/Colour-Word Trial	49	*r* 0.049	95% CI: −0.140, 0.275	*p* .736
		CFI/alternatives subscale	[2]Stroop/Interference Score*	49	*r* 0.063	95% CI: −0.219, 0.366	*p* .666
		CFI/alternatives subscale	COWAT/sum total score	49	*r* 0.114	95% CI: −0.114, 0.324	*p* .436

(*continued on next page*)

Table 2 (*continued*)

Citation & year	Language test(s) were administered in	Self-report test	Neuropsychological test	N in correlation	Raw correlation (*r or rho*)	95% CI or SE	*p*-value
		CFI/alternatives subscale	WCST/perseverative errors	49	*r* −0.214	95% CI: −0.457, 0.062	*p* .141
		CFI/alternatives subscale	WCST/% perseverative responses*	49	*r* −0.208	95% CI: −0.447, 0.060	*p* .151
		CFI/control subscale	TMT/part B time	49	*r* −0.349	95% CI: −0.616, −0.013	*p* .014
		CFI/control subscale	[2]TMT/part B-A time*	49	*r* −0.275	95% CI: −0.587, 0.114	*p* .056
		CFI/control subscale	[2]TMT/part B/A time*	49	*r* −0.143	95% CI: −0.466, 0.209	*p* .328
		CFI/control subscale	Stroop/Colour-Word Trial	49	*r* 0.162	95% CI: −0.059, 0.351	*p* .267
		CFI/control subscale	[2]Stroop/Interference Score*	49	*r* −0.119	95% CI: −0.436, 0.180	*p* .416
		CFI/control subscale	COWAT/sum total score	49	*r* 0.309	95% CI: 0.035, 0.526	*p* .031
		CFI/control subscale	WCST/perseverative errors	49	*r* −0.097	95% CI: −0.356, 0.141	*p* .508
		CFI/control subscale	WCST/% perseverative responses*	49	*r* −0.056	95% CI: −0.331, 0.168	*p* .702
		CFS/total score	TMT/part B time	49	*r* −0.103	95% CI: −0.409, 0.198	*p* .482
		CFS/total score	[2]TMT/part B-A time*	49	*r* −0.001	95% CI: −0.368, 0.315	*p* .995
		CFS/total score	[2]TMT/part B/A time*	49	*r* 0.109	95% CI: −0.255, 0.383	*p* .455
		CFS/total score	Stroop/Colour-Word Trial	49	*r* −0.028	95% CI: −0.308, 0.261	*p* .847
		CFS/total score	[2]Stroop/Interference Score*	49	*r* −0.080	95% CI: −0.327, 0.200	*p* .586
		CFS/total score	COWAT/sum total score	49	*r* 0.113	95% CI: −0.211, 0.382	*p* .437
		CFS/total score	WCST/perseverative errors	49	*r* −0.113	95% CI: −0.428, 0.196	*p* .439
		CFS/total score	WCST/% perseverative responses*	49	*r* −0.120	95% CI: −0.448, 0.201	*p* .413
[1]Kuula et al. (2018)	BRIEF-A, TMT & Stroop; Finnish versions						
Males							
		BRIEF-A/shift subscale*	[2]TMT/part B time	301	*rho* 0.049	95% CI: −0.064, 0.161	*p* .398
		BRIEF-A/shift subscale*	[2]TMT/part B-A time*	301	*rho* 0.014	95% CI: −0.099, 0.126	*p* .810
		BRIEF-A/shift subscale*	[2]TMT/part B/A time*	301	*rho* −0.010	95% CI: −0.122, 0.103	*p* .861
		BRIEF-A/shift subscale*	[2]TMT/part B-A/A time	301	*rho* −0.010	95% CI: −0.122, 0.103	*p* .861
		BRIEF-A/shift subscale*	Stroop Interference Score/part 2 – part 1	301	*rho* −0.025	95% CI: −0.137, 0.088	*p* .666
Females							
		BRIEF-A/shift subscale*	[2]TMT/part B time	386	*rho* 0.143	95% CI: 0.044, 0.239	*p* .005

(*continued on next page*)

Table 2 (*continued*)

Citation & year	Language test(s) were administered in	Self-report test	Neuropsychological test	N in correlation	Raw correlation (*r or rho*)	95% CI or SE	*p*-value
		BRIEF-A/shift subscale*	[2]TMT/ part B-A time*	386	*rho* 0.075	95% CI: −0.025, 0.173	*p* .143
		BRIEF-A/shift subscale*	[2]TMT/part B/A time*	386	*rho* −0.015	95% CI: −0.114, 0.084	*p* .768
		BRIEF-A/shift subscale*	[2]TMT/part B-A/A time	386	*rho* −0.015	95% CI: −0.114, 0.084	*p* .768
		BRIEF-A/shift subscale*	Stroop Interference Score/part 2 – part 1	386	*rho* 0.025	95% CI: −0.075, 0.124	*p* .619
Lau (2012)		CFS/total score	[2]TMT/part B time (t-score)	92	*r* 0.16	95% CI: −0.046, 0.353	ns
		CFS/total score	WCST/perseverative errors (t-score)	92	*r* −0.07	95% CI: −0.27, 0.136	ns
Lounes et al. (2011)		CFS/total score	The Brixton Spatial Anticipation Test/total errors	49	*r* −0.01	95% CI: −0.29, 0.271	*p* > .05
McCloskey (2010)		Questionnaire subtest of Behavioural Rigidity	AUT/forms B and C	75	*r* 0.08	SE 0.11	ns
		Questionnaire subtest of Behavioural Rigidity	Match Problems/form A	75	*r* 0.14	SE 0.11	ns
[1]Meltzer et al. (2017)		BRIEF-A/shift subscale	[2]Stroop/Golden Interference Score*	189	*rho* −0.095	95% CI: −0.236, 0.042	*p* .195
		BRIEF-A/shift subscale	Stroop/Colour-Word Trial	191	*r* −0.007	95% CI: −0.145, 0.136	*p* .928
		BRIEF-A/shift subscale	TMT/part B time	190	*rho* 0.033	95% CI: −0.116, 0.175	*p* .650
		BRIEF-A/shift subscale	[2]TMT/part B-A time*	190	*rho* −0.005	95% CI: −0.150, 0.142	*p* .943
		BRIEF-A/shift subscale	[2]TMT/part B/A time*	190	*rho* −0.061	95% CI: −0.198, 0.074	*p* .403
		BRIEF-A/shift subscale	COWAT/sum total score	194	*r* 0.043	95% CI: −0.097, 0.189	*p* .553
[1]Reed et al. (2020)		BRIEF-A/shift subscale (T1)*	[2]TMT/part A-B time (T1) (scaled score)	141	*rho* 0.205	95% CI: 0.042, 0.371	*p* .015
		BRIEF-A/shift subscale (T1)*	[2]COWAT/sum total score (T1) (scaled score)	147	*r* 0.076	95% CI: −0.084, 0.227	*p* .363
		BRIEF-A/shift subscale (T2)*	[2]TMT/part A-B time (T2) (scaled score)	129	*rho* 0.212	95% CI: 0.042, 0.369	*p* .016
		BRIEF-A/shift subscale (T2)*	[2]COWAT/sum total score (T2) (scaled score)	132	*r* 0.175	95% CI: 0.026, 0.331	*p* .045
		BRIEF-A/shift subscale (T3)*	[2]TMT/part A-B time (T3) (scaled score)	124	*rho* 0.156	95% CI: −0.037, 0.343	*p* .083
		BRIEF-A/shift subscale (T3)*	[2]COWAT/sum total score (T3) (scaled score)	131	*r* 0.219	95% CI: 0.024, 0.409	*p* .012
		BRIEF-A/shift subscale (T4)*	[2]TMT/part A-B time (T4) (scaled score)	120	*rho* 0.130	95% CI: −0.053, 0.286	*p* .158
		BRIEF-A/shift subscale (T4)*	[2]COWAT/sum total score (T4) (scaled score)	128	*r* 0.246	95% CI: 0.101, 0.392	*p* .005
		BRIEF-A/shift subscale (T5)*	[2]TMT/part A-B time (T5) (scaled score)	110	*rho* 0.138	95% CI: −0.063, 0.316	*p* .151
		BRIEF-A/shift subscale (T5)*	[2]COWAT/sum total score (T5) (scaled score)	115	*r* 0.216	95% CI: 0.017, 0.382	*p* .021
		BRIEF-A/shift subscale (T6)*	[2]TMT/part A-B time (T6) (scaled score)	97	*rho* 0.034	95% CI: −0.153, 0.238	*p* .741
		BRIEF-A/shift subscale (T6)*	[2]COWAT/sum total score (T6) (scaled score)	102	*r* 0.185	95% CI: −0.011, 0.396	*p* .063
		BRIEF-A/shift subscale (T7)*	[2]TMT/part A-B time (T7) (scaled score)	80	*rho* 0.232	95% CI: 0.025, 0.434	*p* .038
				88	*r* 0.255		

(*continued on next page*)

Table 2 (*continued*)

Citation & year	Language test(s) were administered in	Self-report test	Neuropsychological test	N in correlation	Raw correlation (*r or rho*)	95% CI or SE	*p*-value
		BRIEF-A/shift subscale (T7)*	[2]COWAT/sum total score (T7) (scaled score)			95% CI: 0.040, 0.448	*p* .016
		BRIEF-A/shift subscale (T8)*	[2]TMT/part A-B time (T8) (scaled score)	72	*rho* −0.222	95% CI: −0.449, 0.013	*p* .061
		BRIEF-A/shift subscale (T8)*	[2]COWAT/sum total score (T8) (scaled score)	75	*r* 0.057	95% CI: −0.156, 0.249	*p* .630
		BRIEF-A/shift subscale (T9)*	[2]TMT/part A-B time (T9) (scaled score)	58	*rho* 0.185	95% CI: −0.076, 0.418	*p* .164
		BRIEF-A/shift subscale (T9)*	[2]COWAT/sum total score (T9) (scaled score)	61	*r* −0.009	95% CI: −0.255, 0.253	*p* .944
		BRIEF-A/shift subscale (T10)*	[2]TMT/part A-B time (T10) (scaled score)	38	*r* −0.156	95% CI: −0.476, 0.192	*p* .349
		BRIEF-A/shift subscale (T10)*	[2]COWAT/sum total score (T10) (scaled score)	41	*r* 0.138	95% CI: −0.140, 0.414	*p* .390
Saylik & Szameitat (2018)		[2]ASQ/positive	[2]CANTAB ID/ED/total errors	90	*r* 0.062	95% CI: −0.146, 0.265	ns
		[2]ASQ/negative	[2]CANTAB ID/ED/total errors	90	*r* −0.016	95% CI: −0.222, 0.191	ns
[1]Zmigrod et al. (2018)		IUT	WCST/% accuracy rate	209	*rho* −0.171	95% CI: −0.300, −0.044	*p* .013
		DOR	WCST/% accuracy rate	209	*rho* −0.049	95% CI: −0.188, 0.095	*p* .481
		IUT	RAT/% accuracy rate	292	*rho* 0.053	95% CI: −0.060, 0.159	*p* .365
		DOR	RAT/% accuracy rate	292	*rho* 0.034	95% CI: −0.091, 0.161	*p* .564

Note. AAQ-II = Acceptance and Action Questionnaire-II; ASQ, = Attributional Style Questionnaire; AUT; Alternate Uses Task; BRIEF-A = Behaviour Rating Inventory of Executive Function-Adult; CANTAB ID/ED = Cambridge Neuropsychological Test Automated Battery Intra-Extra Dimensional Set Shifting Task; CFI = Cognitive Flexibility Inventory; CFS = Cognitive Flexibility Scale; CI = confidence interval; COWAT = Controlled Oral Word Association Test; D-KEFS = Delis-Kaplan Executive Function System; DOR = Dependence on Routines; IRAP = Implicit Relational Assessment Procedure; IUT = Intolerance for Uncertainty; N = sample size; ns = not significant; *p* = *p*-value; *r* = Pearson's correlation coefficient; RAT = Remote Associates Test; *rho* = Spearman's correlation coefficient; SE = standard error; (T1) = time point 1; (T2) = time point 2; (T3) = time point 3; (T4) = time point 4; (T5) = time point 5; (T6) = time point 6; (T7) = time point 7; (T8) = time point 8; (T9) = time point 9; (T10) = time point 10; TMT = Trail Making Test; WCST = Wisconsin Card Sorting Test; * = additional outcomes that were obtained from the authors (i.e., not reported in the original citation); (^) = correlation type (Pearson's or Spearman's) was not reported or could not be confirmed with the authors; [1] = reported data does not match the data reported in the original article in order to fulfill the eligibility criteria; [2] = direction of test scores (i.e., higher/lower scores indicate better/worse performance) could not be established.

(COWAT; Benton, Hamsher, & Sivan, 1994), two studies (Aytug et al., 2018; McCloskey, 2010) used the Alternate Uses Task (Guilford, 1956), and two studies (Bellinger et al., 2015; Dawson et al., 2012) used the Delis-Kaplan Executive Function System (Delis et al., 2004; Delis, Kaplan, Kramer, & Functioning, 2001) battery, specifically the Verbal Fluency Task and the Colour-Word Interference Task, respectively. One study (McCloskey, 2010) used the Match Problems task (Guilford & Guilford, 1980), one study (Lounes et al., 2011) used the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997), one study (Cohen et al., 2017) used the Set-shifting task (Kramer et al., 2014), one study (Saylik & Szameitat, 2018) used the Intra-Extra Dimensional Set Shift Task of the Cambridge Neuropsychological Test Automated Battery (Robbins et al., 1998), one study (Drake, 2008) used the Implicit Relational Assessment Procedure, and one study (Zmigrod et al., 2018) used the Remote Associates Test (Mednick, 1968). Throughout the included studies, different versions of the WCST, TMT, Stroop and COWAT were used. In some instances, it was difficult to determine which version of the task had been used (i.e., manual or computerised version, original or subsequent iterations). The languages in which the tests were administered were often not reported but were verified via author contact.

(i) *Fig. 4: Correlations of all test pairs accounting for dependency of effect sizes (multi-level random-effects meta-analysis)*: One hundred and nineteen comparisons were obtained from across 21 studies. Seventeen comparisons (14%) showed a significant relationship in the predicted direction; eight comparisons (7%) showed a significant relationship in the direction opposite to that predicted; 94 comparisons (79%) were non-significant. The pooled correlation was 0.05 (95% CI 0.00 to 0.10), demonstrating that there was no relationship between self-report and neuropsychological tests of cognitive flexibility (see Fig. 4).

(ii) *Level of variance within the multi-level random-effects model*: We estimated the variance being contributed to each level of the model using a likelihood-ratio test and found that when we compared the full model to a reduced model (i.e., a two-level model without within-study variance), the within-study

Table 3
Reported definitions of cognitive flexibility for included studies.

Definition of cognitive flexibility	Citation(s) & year
"Executive functions (EF), also referred to as cognitive control, comprise processes that are necessary for goal-oriented behavior. They include at least three broad categories of basic executive processes: working memory/maintenance, response inhibition and cognitive flexibility, and higher-order EF (e.g., planning) built upon combinations of these three components (for review see Diamond, 2013; Etkin, Gyurak, & O'Hara, 2013)".	Alfimova et al. (2017)
"Cognitive flexibility refers to the ability to adapt cognitive processing strategies in the face of new and unexpected conditions in one's environment (Canas, Quesada, Antolí, & Fajardo, 2003). It reflects (i) an awareness of available communication alternatives, (ii) a willingness to adapt to the present situation, and (iii) a self-efficacy in being flexible (Martin & Anderson, 1998; Martin & Rubin, 1995). Such cognitive flexibility allows individuals to efficiently switch between different behaviors and strategies when faced with novel situations and environmental demands (Gocłowska, Crisp, & Labuschagne, 2012)".	Aytug et al. (2018)
"Cognitive flexibility has been defined as a part of executive functioning that refers to the ability to shift cognitive sets, thought, or attention, to perceive, process, or respond to situations in different ways, such as by generating multiple ideas, switching between different types of knowledge, inhibiting habitual responses in favor of alternative responses, and modifying behavior and cognition when required by changing environmental circumstances (Eslinger & Grattan, 1993; Rende, 2000; Johnco, Wuthrich, & Rapee, 2013)".	Brenner (2017)
• "[…] cognitive flexibility, which has been defined as the ability to successfully integrate and apply previously learned information (Kanagaratnam & Asbjørnsen 2007)".	Clayson (2018)
• "[…] set shifting (also referred to as *cognitive flexibility*) refers to a flexible information processing style characterized by successful integration and application of previously learned information (Kanagaratnam & Asbjørnsen, 2007)".	
"Executive function is an umbrella term that refers to three main cognitive functions: (1) working memory—maintaining and updating relevant information; (2) inhibition—inhibiting prepotent impulses, and (3) shifting—rapidly and efficiently adapting to different situations (Miyake et al., 2000)".	Dassen et al. (2018)
"[…] *psychological flexibility*, defined as "the ability to contact the present moment more fully as a conscious human being, and to either change or persist when doing so serves values ends" (Hayes, Luoma, Bond, Masuda, & Lillis, 2006, p.5)".	Drake (2008)
"Cognitive flexibility has been defined as the ability to shift directions of thought and action to perceive, process and respond to situations flexibly and adaptively (Eslinger & Grattan, 1993)".	Hertzberg (2017) (Sample B)
"Cognitive flexibility (CF) is one component of executive functioning that refers to the ability to freely shift cognitive sets to perceive or respond to situations in different ways, such as by generating multiple ideas, switching between different classes of knowledge, and inhibiting habitual responses in favor of alternative responses when required by changing environmental circumstances (Rende, 2000)".	Johnco et al. (2014)
"Executive functions (EFs) are considered the central control systems for complex behaviour, such as attention, planning and goal-setting, inhibitory abilities and cognitive flexibility".	Kuula et al. (2018)
"Cognitive flexibility was conceptualized as an objective measurement of executive processes (e.g., mental flexibility and set-shifting) and a subjective perception of self-efficacy".	Lau (2012)
"In particular, set-shifting, a major component of cognitive flexibility (Tchanturia et al., 2005), involves the ability to disengage from an irrelevant task set and actively engage with a relevant task set".	Lounes et al. (2011)
"[…] spontaneous flexibility as "the ability or disposition to produce a diversity of ideas when free to do so" (Guilford, 1959, p. 383), and adaptive flexibility as the ability to change set and to restructure or redefine problems to solve them (Guilford & Guilford, 1980)".	McCloskey (2010)
"First, executive function (EF) reflects the ability to perform set-shifting, inhibition, and updating and to utilize working memory effectively, typically acting on abstract content such as numbers, letters, and colors and measured using neuropsychological tests (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000)".	Reed et al. (2020)
"While switching allows the shifting of one's attention from one task to another […]"	Saylik & Szameitat (2018)
No definition of cognitive flexibility reported	Bellinger et al. (2015); Cohen et al. (2017); Dawson et al. (2012); Dingemans et al. (2019); Hertzberg (2017) (Sample A); Meltzer et al. (2017); Zmigrod et al. (2018)

variance was non-significant (p .5). That is, we found no significant variability between effect-sizes within studies. We then conducted another likelihood-ratio test by comparing the full model to a reduced model (i.e., a two-level model without between-study variance), and found that between-study variance was significant (p <.0001). That is, we found significant variability between studies. We were then able to calculate the distribution of the total variance across the three levels of the random effects model and found that 47.2% of the variance is explained by sampling variance, 4.4% is between effect sizes within each study, and 48.5% is between-study variance.

(iii) *Moderator analysis*: Our moderator analysis found no difference in the effects by age group (younger adults versus older adults). That is, the model revealed a non-significant effect of 0.04 (95% CI 0.17 to 0.10).

(iv) *Sensitivity analysis*: Our sensitivity analysis, removing studies that provided a conflict of interest statement, revealed a significant effect of 0.09 (95% CI 0.01 to 0.17). That is, there was a small but significant relationship between self-report and neuropsychological tests of cognitive flexibility.

Table 4
Risk of bias assessment.

Study	Were the aim(s) and/or research question(s) stated, and aligned with our research question?	Were the sample demographics (i.e., age, gender/sex) described/ obtained?	Was the sample size justified for the correlational analyses?	Do the authors report the tests as reliable?	Were the study conditions implemented consistently across all participants?	Were the descriptive data of the tests described/ obtained?	Were the results of the correlation described/ obtained?	Were planned correlational analyses performed and stated a priori?	Were the relevant data internally consistent?	If parametric tests were used, were the assumptions of normality met for the correlation?	Was the test order counterbalanced across participants and planned for in the analyses?	Were funding sources declared?	Was a conflict of interest statement provided?	Was a protocol registered a priori?
Alfimova et al. (2017)	N	Y	N	N	?	Y	Y	N/A	N	Y	?	N	N	N
Aytug et al. (2018)	N	Y	Y	?	?	Y	N	?	Y	?	?	N	N	N
Bellinger et al. (2015)	N	Y	N	N	?	Y	Y	N/A	Y	Y	?	Y	Y	N
Brenner (2017)	Y	Y	N	Y	Y	Y	Y	?	N	Y	?	N	N	N
Clayson (2018)	N	Y	N	?	Y	Y	Y	N/A	Y	Y	N	N	N	N
Cohen et al. (2017)	N	Y	N	?	Y	Y	Y	N/A	Y	Y	?	Y	Y	N
Dassen et al. (2018)	N	Y	N	N	N	Y	Y	N/A	Y	Y	N	Y	Y	N
Dawson et al. (2012)	N	Y	N	N	N	Y	Y	N/A	Y	Y	N	Y	N	N
Dingemans et al. (2019)	N	Y	N	?	?	Y	Y	N/A	Y	?	?	N	Y	N
Drake (2008)	N	Y	N	?	Y	Y	Y	?	Y	N	Y	N	N	N
Hertzberg (2017) (Sample A)	N	Y	N	N	?	Y	N	?	Y	?	?	N	N	N
Hertzberg (2017) (Sample B)	Y	Y	N	?	?	Y	Y	Y	Y	?	?	N	N	N
Johnco et al. (2014)	Y	Y	N	?	?	Y	Y	?	Y	Y	?	N	N	N
Kuula et al. (2018)	N	Y	N	?	?	Y	Y	N/A	Y	N/A	?	N	Y	N
Lau (2012)	N	Y	N	?	N	N	Y	Y	N	Y	N	N	N	N
Lounes et al. (2011)	Y	Y	N	?	?	Y	Y	?	Y	Y	?	Y	Y	N
McCloskey et al. (2010)	N	Y	N	Y	?	Y	Y	?	Y	Y	N	N	N	N
Meltzer et al. (2017)	N	Y	N	?	Y	Y	Y	N/A	Y	Y	N	Y	Y	N
Reed et al. (2020)	N	Y	N	?	?	Y	Y	N/A	Y	N/A	?	Y	Y	N
Saylik & Szameitat (2018)	N	Y	N	N	Y	N	Y	?	N	Y	N	N	Y	N
Zmigrod et al. (2018)	N	Y	N	?	Y	Y	Y	?	N	N/A	N	Y	Y	N

Note. Y = Yes; N = No; ? = Indeterminate; N/A = not applicable; Red = high risk of bias; Green = low risk of bias; Yellow = Indeterminate risk of bias; Blue = not applicable.

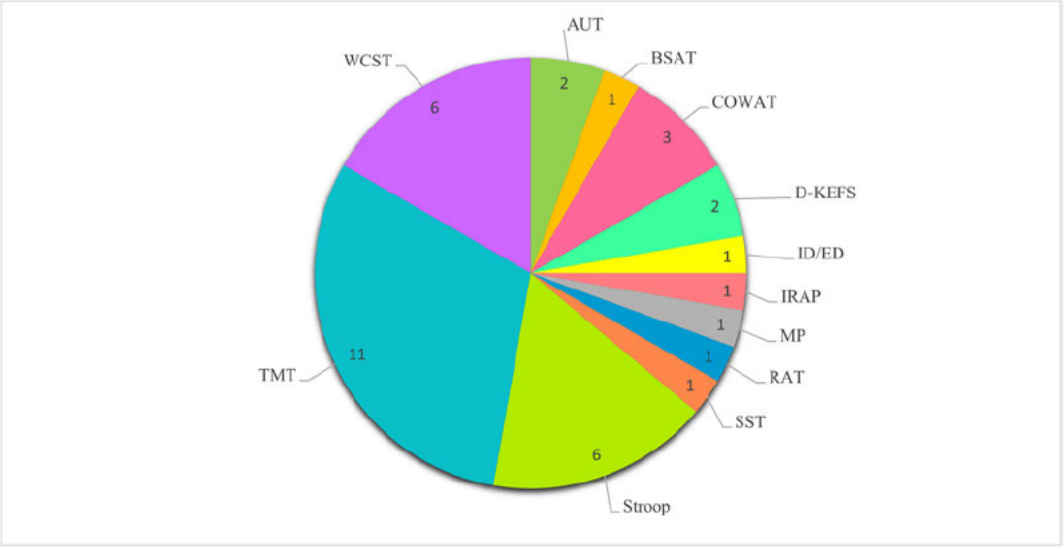


Fig. 3. Number of included studies that used each neuropsychological test of cognitive flexibility.
Note. AUT = Alternate Uses Task; BSAT = Brixton Spatial Anticipation Test; COWAT; Controlled Oral Word Association Test; D-KEFS; Delis-Kaplan Executive Function System; ID/ED = Intra-Extra Dimensional Set Shift Task; IRAP = Implicit Relational Assessment Procedure; MP = Match Problems task; RAT = Remote Associates Test; SST = Set-shifting task; Stroop = Stroop test; TMT = TMT = Trail Making Test; WCST; Wisconsin Card Sorting Test.

(v) *Fig. 5 and Fig. 6: Random-effects meta-analyses between comparisons of the same test pair(s) from different cohorts*: The most common comparison was between the shift subscale of the BRIEF-A and the TMT – part B (time) (see Fig. 5). The pooled correlation from eight comparisons showed no relationship (r 0.08, 95% CI 0.02 to 0.19). Statistical and substantial heterogeneity was detected for this test pair (Q 15.73, p .03, I^2 61.0%). There were ten other comparisons that were undertaken on multiple cohorts. The only significant relationship was that between the CFS and the TMT – part B (time) (see Fig. 6), effect estimate 0.19 (95% CI 0.06 to 0.31), thus we constructed a third forest plot. Forest plots for the remaining nine comparisons are shown in the Supplementary Files (see Supplementary File S3).

3. Discussion

3.1. Summary and interpretation of findings

The aim of the current systematic review was to determine whether, and to what extent, self-report and neuropsychological tests of cognitive flexibility are related in healthy adults, and thus substantiate the implicit assumption that the two assessment approaches provide different windows into the same room of cognitive flexibility. We found no relationship, which suggests that greater care should be taken when planning and reporting experiments and clinical studies, and that self-report questionnaires should not be considered as valid proxies for neuropsychological tests of cognitive flexibility. Further, the extrapolation from neuropsychological test scores to observed or reported behavioural rigidity should be avoided.

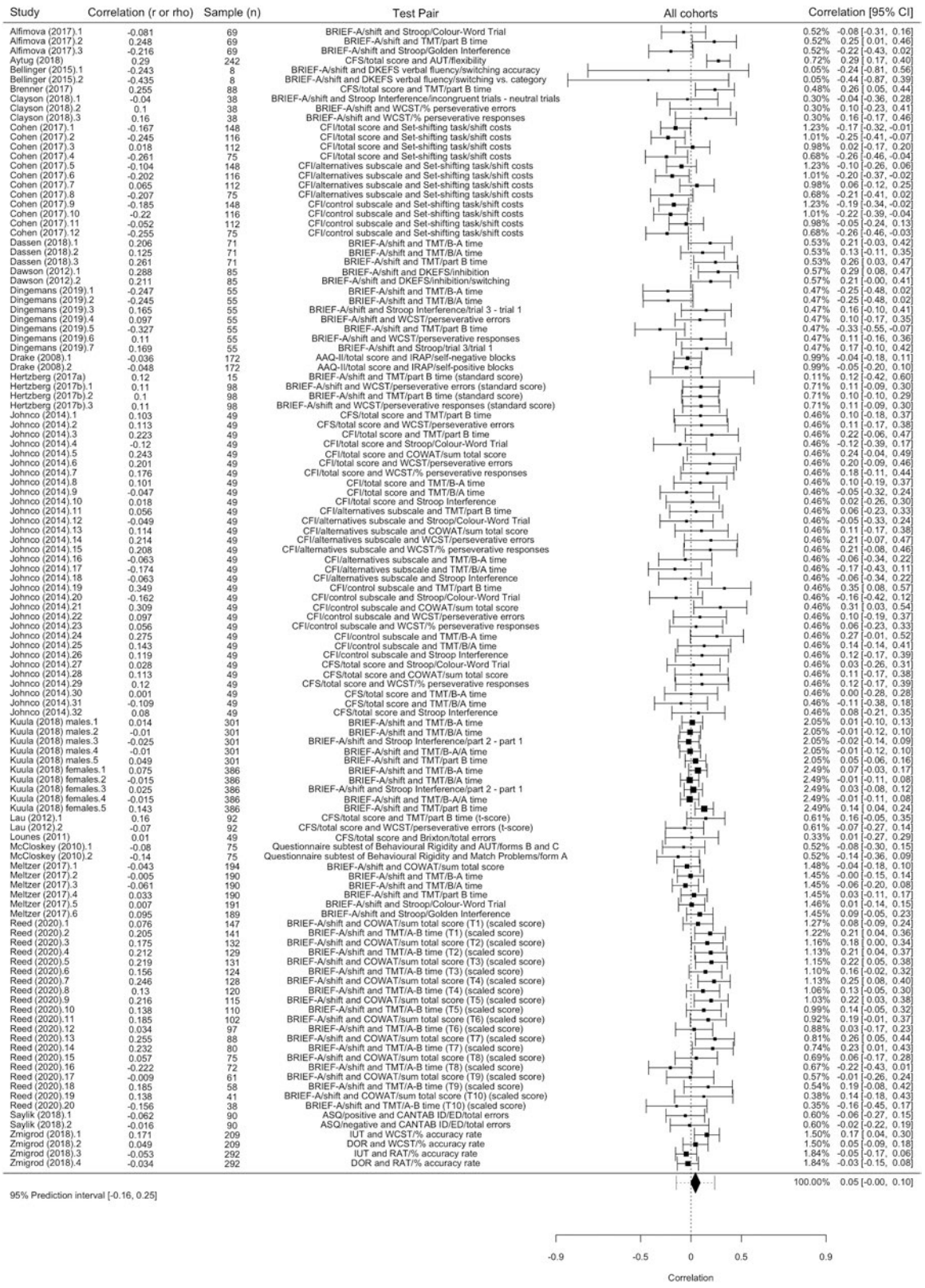


Fig. 4. Correlations of all test pairs accounting for dependency of effect sizes (multi-level random-effects meta-analysis).

Note. AAQ-II = Acceptance and Action Questionnaire-II; ASQ = Attributional Style Questionnaire; AUT = Alternate Uses Task; BRIEF-A = Behaviour Rating Inventory of Executive Function-Adult; Brixton = The Brixton Spatial Anticipation Test; CANTAB ID/ED = Cambridge Neuropsychological Test Automated Battery Intra-Extra Dimensional Set Shifting Task; CFI = Cognitive Flexibility Inventory; CFS = Cognitive Flexibility Scale; CI = confidence interval; COWAT = Controlled Oral Word Association Test; D-KEFS = Delis-Kaplan Executive Function System; DOR = Dependence on Routines; IRAP = Implicit Relational Assessment Procedure; IUT = Intolerance for Uncertainty; n = sample size; *r* = Pearson's correlation coefficient; RAT = Remote Associates Test; *rho* = Spearman's correlation coefficient; TMT = Trail Making Test; WCST = Wisconsin Card Sorting Test.

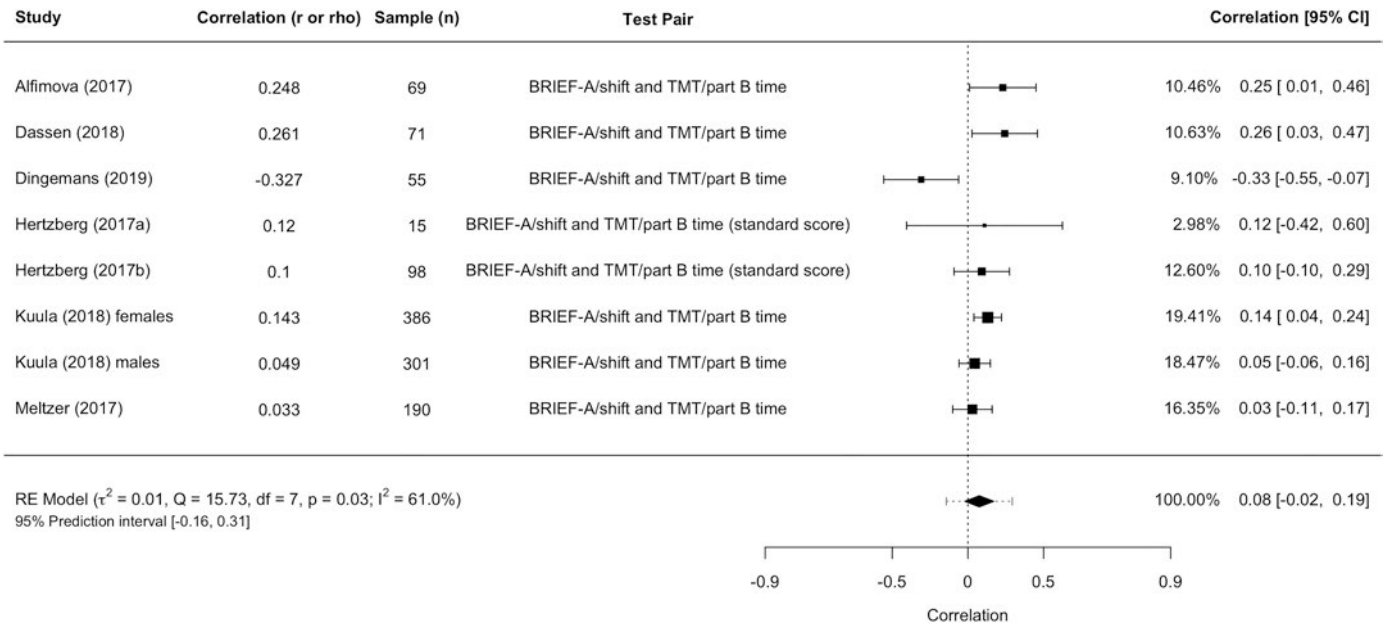


Fig. 5. Meta-analysis of pooled cohorts that used the shift subscale of the BRIEF-A and the TMT – part B (time).
Note. BRIEF-A = BehaviourRating Inventory of Executive Function-Adult; CI = confidence interval; n = sample size; *r* = Pearson's correlation coefficient; *rho* = Spearman's correlation coefficient; TMT = Trail Making Test.

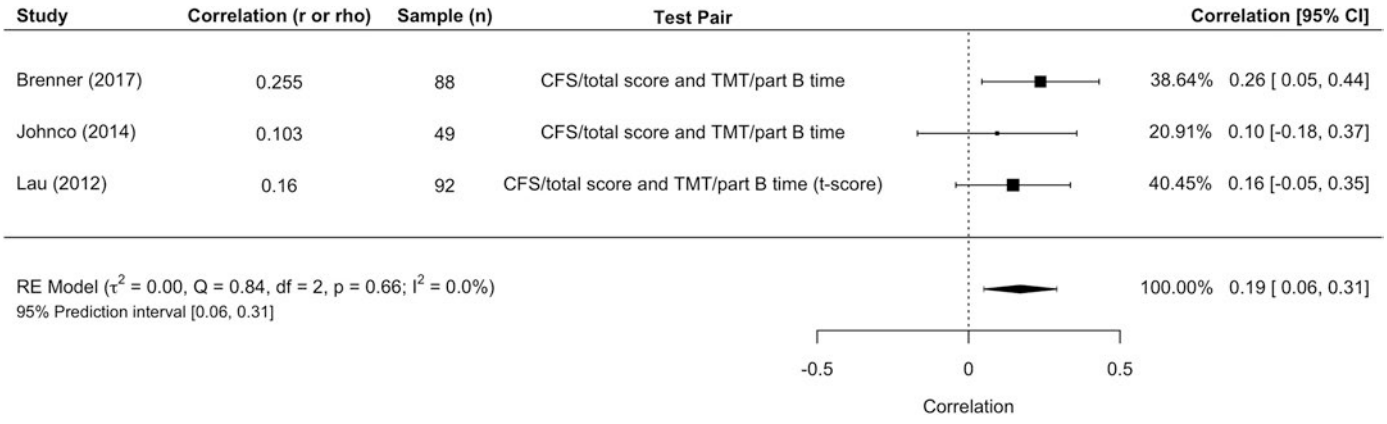


Fig. 6. Meta-analysis of pooled cohorts that used the CFS and the TMT – part B (time).
Note. CFS = Cognitive Flexibility Scale; CI = confidence interval; n = sample size; *r* = Pearson's correlation coefficient; *rho* = Spearman's correlation coefficient; TMT = Trail Making Test.

A multi-level random-effects meta-analysis revealed no relationship. Upon investigating the amount of variance being contributed to each level of the multi-level random-effects model, we detected significant variability between studies, but this heterogeneity could not be explained by differences in the effects by age group. Individual meta-analyses between test pairs revealed that the CFS and the TMT – part B (time) was the only test pair to demonstrate a significant relationship, although it was small and whether or not it would be considered meaningful is up for debate. All of the remaining test pairs were unrelated. Considering the threats to validity of several studies, we conclude that there is likely to be no meaningful relationship between self-report and neuropsychological tests of cognitive flexibility, a situation that has broad ramifications for our field.

Our findings do not support our original prediction, which was based on anecdotal and inferred evidence that the assumption that self-report and neuropsychological assessments of cognitive flexibility assess the same construct. Some of the definitions used in the primary data here (see Table 3), for example Lau (2012), Aytug et al. (2018), and Hertzberg (2017) (Sample B), clearly show this assumption. Surprisingly, we are among the first to ask this question and the first to seek out all available data in order to answer it. We are not the first to observe the lack of relationship however: Lounes et al. (2011), Gelonch, Garolera, Valls, Rosselló, and Pifarré (2016), Hertzberg (2017) (Sample A and B) and Gonzalez et al. (2013) reported no relationship between self-report and neuropsychological tests of cognitive flexibility in any of their samples: females with and without anorexia nervosa, people with and without fibromyalgia, alcohol-abusing individuals, and undergraduate students, respectively. Johnco et al. (2014) found a small relationship between some self-report and neuropsychological tests of cognitive flexibility but no relationship for others, in their non-clinical sample of older adults. Additionally, Brenner (2017) found a small relationship between the CFS and the TMT – part B (time) in male perpetrators of domestic violence, which is the same test pair that produced a small relationship in our meta-analysis. Critically, however, Brenner (2017)

identified a negative relationship between this test pair whereas our meta-analysis identified a positive relationship. Among the aforementioned studies, Brenner (2017), was the only one to report power calculations, although they did not reach their required sample. Perhaps the field has not questioned the assumption because the studies that did in fact test it were underpowered to detect the relationship anyway. This possibility underpins the need for our broad-based systematic search and review. A lack of relationship between self-report and neuropsychological tests has been reported in other domains of executive function (Burmester, Leathem, & Merrick, 2016; Rabin et al., 2006; Spencer, Drag, Walker, & Bieliauskas, 2010), which demonstrates that the problem is clearly not confined to the interrogation of cognitive flexibility.

Our results cement our initial observations of a prevalent problem in the field – the lack of consensus around defining and describing cognitive flexibility (see Table 3). Eight studies stated broadly similar (but not identical) definitions along the lines of "the ability to shift mental sets" or "adapt cognitive strategies" (Aytug et al., 2018; Brenner, 2017; Hertzberg, 2017 (Sample B); Johnco et al., 2014; Lau, 2012; Lounes et al., 2011; McCloskey, 2010; Saylik & Szameitat, 2018), but there were several other definitions used and seven studies reported no definition at all. One study used the term *psychological flexibility* and defined it in line with the ACT theory as the ability to remain in the present moment, and either persist or adapt behaviour when doing so is coherent with one's values and/or goals (Drake, 2008). Three studies (Alfimova et al., 2017; Meltzer et al., 2017; Reed et al., 2020) used the terms *cognitive flexibility* and *set-shifting* interchangeably and five studies (Brenner, 2017; Johnco et al., 2014; Kuula et al., 2018; Lau, 2012; Reed et al., 2020) used the terms *cognitive* and *mental* flexibility synonymously. There are other instances throughout the literature where the terms *cognitive, mental, psychological* and *set-shifting* have been used interchangeably (Davis, Marra, Najafzadeh, & Liu-Ambrose, 2010; Diamond, 2013; Leung & Zakzanis, 2014; Palm & Follette, 2011; Shepherd, Reynolds, Turner, O'Boyle, & Thompson, 2019; Wecker et al., 2005). This is clearly problematic for our field – if consensus in defining and delineating these constructs cannot be established, operationalising them will also remain elusive.

3.2. Risk of bias

All 21 included studies were evaluated as having a moderate to high risk of bias. Contributors to this risk of bias came both from the studies themselves – most commonly methodological and reporting problems – and that our question was not the primary question for 17 out of 21 included studies. Notably, we judged papers whose research aims did not align with ours as *high risk of bias* because we expected those papers to be less precise about factors deemed important to our question than they were for factors important to their question; however, it has been argued that including empirical studies that do not share identical research goals to meta-analyses may be less vulnerable to publication bias because publication success is not determined by factors related to the question itself (Lange, 2020). Of particular note was that no study pre-registered a protocol, which is important for transparent reporting and replicability and is broadly recommended within psychological sciences (Klein et al., 2018) and related fields (e.g., Lee et al., 2018), and only two studies (Brenner, 2017; McCloskey, 2010) reported the reliability of the tests they used. The widespread lack of rigour and transparency points to clear opportunities to improve the evidence base and our understanding of cognitive flexibility moving forward.

3.3. Strengths, limitations and considerations

This systematic review has several noteworthy strengths and limitations. The protocol for this review was pre-registered on PROSPERO (registration number: CRD42019131567; available from https://www.crd.york.ac.uk/prospero/display_record.php?ID CRD42019131567), as well as Open Science Framework (https://osf.io/7yzmq/), and any deviations have been fully disclosed and time-stamped. The PRISMA statement (Moher et al., 2009) was adhered to throughout the entire review process, and two independent reviewers completed article screening, data extraction, risk of bias assessments and data analysis. The search strategy was very comprehensive. We did not limit the search in terms of language, country of origin, date of publication or publication status. However, we chose not to include the term *executive function* in our search strategy because of its broad nature, and our preliminary searches illustrated that it would be an unmanageable process. We cannot exclude the possibility that our approach may have resulted in papers being missed; however, the large yield leaves us confident that anything we did miss would not have changed the result. We went to great lengths to contact authors via email, phone and ResearchGate (ResearchGate GmbH, Berlin, Germany), and were thus able to obtain raw or aggregate data and confirm inclusion criteria with 20 authors. Despite our attempts, however, we were unable to include the majority of studies we identified because we could not contact authors, or the authors were unable to provide the required data. Even when we could, important aspects of some studies still could not be established, for example, whether Pearson's *r* or Spearman's rho were used, or the direction of the test scores (i.e., whether higher or lower scores indicate better or worse performance). We had not planned for that scenario in our *a priori* protocol, so we inferred the direction of the test scores according to what would be predicted if they were related. That some reports had correlations in the opposite direction to that predicted undermines our inference, but that all reports had correlations crossing or very close to zero suggests that any impact on the findings would be trivial. Notably, we deviated from our *a priori* protocol when we updated the search strategy in October 2020. We re-ran the database searches but not the grey literature searches. We made this choice because determining the eligibility of data from grey literature 'hits' was very time consuming. In addition, obtaining data from eligible studies was also very time consuming. Therefore, before re-running the search, we decided that additional studies that were identified through our updated search were included providing the data required for inclusion were reported in the published paper. That our updated search did not result in the inclusion of any additional papers does raise the possibility that very recent data have been missed because of this strategy. For example, a recent paper by Lange and Dewitte (2019), which investigated cognitive flexibility and pro-environmental behaviour – and performed correlations between various self-report and neuropsychological tests of cognitive flexibility – commendably provided a link to data that may well have been relevant. Importantly, however, the sample size involved, and the precision of the effect estimates obtained in this review, means that such data would be highly unlikely to affect our main findings.

We chose to include studies regardless of their aim or research question. That our research question was not the primary question for most studies included in this review points to the comprehensive nature of our search, but also raises the possibility that even our pooled estimates remain underpowered to detect a relationship. We suspect that this remains unlikely however, because the forest plots clearly reveal study-specific results either side of zero correlation. Moreover, even if an effect does in fact exist in healthy adults, it is likely to be too small to support the assumption that the two assessment approaches interrogate the same construct.

One novel aspect of our protocol was the *a priori* plan to undertake a sensitivity analysis based on the declaration or not of conflicts of interest. The potential bias introduced by conflict of interests is well recognised (Resnik, 2007), but its effects are seldom considered in systematic reviews. We found that when the studies that included a conflict of interest statement were removed, there was a small but significant relationship between self-report and neuropsychological assessments, which might suggest merit in our intent. We did not formally assess publication bias of the included studies, but we think publication

bias was unlikely because: one would predict that the result of a publication bias would be an overestimation of the relationship between self-report and neuropsychological tests of cognitive flexibility, and we found no relationship; and second, most of our data were obtained from studies that were not asking the same question as ours, so publication bias in those studies would not be relevant to our question.

Most studies used different combinations of tests, making comparisons between studies difficult. Meta-analyses were possible when (i) multiple studies either used the same outcome variables, (ii) we were provided with additional outcome variables for the same test from authors or (iii) we calculated these outcome variables ourselves from the raw data. Even when independent studies used the same test, there was substantial variability in the outcome measures that were drawn from that test. This was particularly the case for the TMT and the WCST. To counteract this issue in our review, we requested multiple outcome measures for these tests (e.g., TMT – part B (time) and TMT B-A (time)) or calculated these outcome variables from the raw data ourselves to allow for the meta-analyses presented in the Supplementary Files (see Supplementary File S3). We are not the first in this field to encounter this problem – previous systematic reviews have observed a diverse range of assessments and outcome variables used to assess cognitive flexibility (Berryman et al., 2014; Bunk et al., 2019; Miles et al., 2020; Roberts et al., 2007). Clearly, this presents challenges because it makes pooling across studies difficult (Lang, Stahl, Espie, Treasure, & Tchanturia, 2014). Another issue that was apparent from this review was how TMT data are reported. Part B involves both psychomotor speed and flexibility components, whereas part A involves just psychomotor speed. The vast majority of the studies included in this review either reported part B alone or reported part A and part B separately. Variability in psychomotor speed introduces 'noise' into part B data (Vall & Wade, 2015) and reduces the likelihood of detecting the relationships we were interested in here; therefore, controlling for part A when reporting the data for part B is most likely to reflect the flexibility component.

Finally, we acknowledge that many of the neuropsychological tests covered here are not thought to uniquely assess cognitive flexibility (Miyake et al., 2000). Because of the lack of clarity in the field about assessing cognitive flexibility, we chose to include any paper that either used a test that had previously been used to assess cognitive flexibility (determined *a priori*), or in which the authors stated *a priori* that the test they were using was an assessment of cognitive flexibility. This approach put the onus on the authors of primary studies to determine whether or not the tests they used were assessing cognitive flexibility, which may have resulted in the inclusion of imprecise assessments and the exclusion of precise assessments.

3.4. Recommendations for future research

The current findings have implications for future research in healthy adults. The first priority should be to reconsider how we use, or in fact define, the term 'cognitive flexibility' (or synonymously used terms). Whether or not we should persevere with the broad construct, or adopt more precise terms that fit the assessments that are used, seems worthy of debate (although, notably, the extensive data reviewed here are limited to healthy adult volunteers and it remains possible that data obtained from clinical populations may detect a relationship that was not detected here). We strongly suggest that expert consensus, perhaps through a Delphi-style study, will be an important step toward more clarity and utility around the broad construct of 'cognitive flexibility'. Another important next step would be to verify the current findings in clinical cohorts; and thus, a similarly comprehensive systematic review is clearly required.

Once conventions to promote greater clarity and precision have been achieved, more theoretical and empirical research could then establish whether, and how, various tests interrogate various sub-types of cognitive flexibility and how they relate to each other. In the meantime, it would appear prudent to report on self-report assessments of cognitive flexibility in a manner that reflects the self-rated ability to adapt to new and unexpected conditions in one's environment, and neuropsychological tests in a manner that reflects the actual test undertaken. We also propose that the term 'psychological flexibility' be confined to the discussion of ACT because it was first coined within that context, is predominantly (although not exclusively) used with reference to ACT, and is usually defined in clinical terms. The original definition – "the ability to contact the present moment more fully as a conscious human being, and to either change or persist when doing so serves values ends" (Hayes et al., 2006, p. 7) – is partly metaphorical and therefore not well suited to empirical quantitative interrogation. Finally, we propose that synonymous labels such as *mental flexibility* be avoided because they ultimately create confusion for researchers and clinicians alike.

Future exploration of the lack of relationship between self-report and neuropsychological assessments of cognitive flexibility, and perhaps executive function in general, might address the *task impurity problem* – a notorious limitation of neuropsychological tests (Miyake et al., 2000). The task impurity problem is the issue that neuropsychological assessments do not exclusively assess one component of executive function (i.e., cognitive flexibility); and thus to successfully complete a neuropsychological test of cognitive flexibility requires both executive and non-executive capacities that extend beyond just cognitive flexibility, or other aspects of executive function, and includes an array of processes that rely on distributed brain activity (Buchsbaum, Greer, Chang, & Berman, 2005; Miyake et al., 2000; Wildes, Forbes, & Marcus, 2014). Task impurity also clouds assessments of self-reported cognitive flexibility; they require capacities beyond cognitive flexibility, including self-awareness. Importantly, the extra capacities required for neuropsychological tests are different from those required for self-report tests.

Our risk of bias assessments highlighted common methodological problems that can be readily overcome. Our recommendations mirror those that have been made before in related fields (e.g., Lee et al., 2018): (i) studies should pre-register their protocol *a priori,* declare funding and conflicts of interest so as to provide higher standards of transparency, accountability and thorough reporting, and declare the language in which tests have been administered, (ii) sample size should be calculated *a priori,* based on the effect to be detected and cater for multiple outcomes and analyses, (iii) when various tests are being administered, the potential of one test to systematically affect another should be mitigated by, for example, extending the delay between the two tests, adding filler tasks between them, or counterbalancing the order of the tests across participants, (iv) protocols should include a statistical analysis plan, or a separate statistical analysis plan should be lodged prior to commencing analysis, (v) studies should either (a) report the reliability and validity of the tests being used (with citations), (b) validate the tests in the population of interest or (c) declare when reliability and validity are not established because of a lack of research, or because the task was created by the study authors.

Other findings of this review have important methodological implications. We recommend that researchers report which scoring method of a test was used (i.e., Grant & Berg, 1948 or Heaton et al., 1993 for the WCST – see also Miles et al., 2021), and the modality of test administration (i.e., manual or computerised administration). Future research using the TMT should control for baseline motor speed because failing to do so has the potential to cause misinterpretations of findings. Further, where the findings of part A and part B are reported separately, researchers and clinicians should be mindful of the problems with reporting TMT – part B in isolation and take the results for TMT – part A into consideration when interpreting the findings. Finally, it is important that clinicians and researchers consider that neuropsychological tests targeting cognitive flexibility quantify processing speed during tasks with changing or alternative rules while self-report questionnaires of cognitive flexibility rely on one's self-appraisal of how they think and behave. As such, the former should not be used for assessing the latter and vice versa.

4. Conclusion

The aim of this systematic review and meta-analysis was to determine whether, and to what extent, self-report and neuropsychological tests of cognitive flexibility are related in healthy adults. Our results do not support the assumption that the two assessment approaches provide different windows into the same broad room of cognitive flexibility in healthy adults. This review captures the inconsistencies within the field with respect to defining and assessing cognitive flexibility, and widespread methodological and reporting problems. We advise against the interchangeable use or generalisation of self-report and neuropsychological tests of cognitive flexibility, and urge the field to seek definitional consensus and revisit what the various tests purported to interrogate cognitive flexibility actually reflect.

Role of funding sources

CAH and SM were recipients of Australian Government Research Training Program (RTP) scholarships. CB was supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship [ID 1127155]. MAW was supported by a Postgraduate Scholarship from the NHMRC, a School of Medical Sciences Top-Up Scholarship from the University of New South Wales, and a PhD Top-Up Scholarship from Neuroscience Research Australia. GLM and EM were supported by a NHMRC Leadership Investigator Grant [ID 1178444] to GLM. EK was supported by a project grant from Lifetime Support Authority of South Australia. These funding bodies had no role in the study design, collection, analysis or interpretation of the data, writing of the manuscript, or the decision to submit the manuscript for publication.

Contributors

Authors CAH, EK, CB and GLM designed the systematic review. CAH and GLM wrote and pre-registered the protocol. CAH conducted the literature search and the following authors: CAH, LP, EJ, SM, AO, EM, KS screened all articles. CAH [all articles], LP, AO, EM and KS conducted data extraction and risk of bias assessments. CAH, KS and MAW conducted data analysis. CAH wrote the first draft of the manuscript and all authors contributed to and have approved the final version of the manuscript.

Declaration of Competing Interest

GLM has received support from: Reality Health, ConnectHealth UK, Seqirus, Kaiser Permanente, Workers' Compensation Boards in Australia, Europe and North America, AIA Australia, the International Olympic Committee, Port Adelaide Football Club, Arsenal Football Club. Professional and scientific bodies have reimbursed him for travel costs related to presentation of research at scientific conferences/symposia. He has received speaker fees for lectures on pain and rehabilitation. He receives book royalties from NOIgroup publications, Dancing Giraffe Press & OPTP. CB has received support from Workers' Compensation Boards in Australia. All remaining authors declare no potential conflicts of interest (personal or financial) with respect to the research, authorship, and/or publication of this manuscript.

Acknowledgements

We would like to thank all of the authors who responded to us to confirm study eligibility. We would also like to sincerely thank all the authors of included studies who responded to our requests for data and/or provided clarification. The level of generosity and assistance with our review is appreciated, and this review would not have been possible without their contribution. We also wish to thank two anonymous reviewers for their thoughtful comments and helpful suggestions on the earlier version of the manuscript.

Appendix A. Medline search strategy

Database: Ovid MEDLINE (R) <1948 to May 10, 2019>
Search Strategy:

1. stroop test/ (1526)
2. trail making test/ (770)
3. wisconsin card sorting test/ (25)
4. (brixton task or brixton spatial anticipation task or catbat task or haptic illusion task or haptic illusion* or "set? shifting subset of cantab" or cantab or cambridge neuropsychological test? or cambridge neuropsychological test automated battery or intra?extra dimensional set shift* or intra dimensional shift* test or extra dimensional shift* test or IDED or IED or multi?task* test? or MTT or delis-kaplan executive function* or delis kaplan executive function* or stroop or colo?r?word interference test? or halstead* or seashore rhythm? or m-wcst or wisconsin card? sort* or modified wisconsin card sort* test or alternat* use? task? or trail mak* or verbal fluency test or word context test).mp. (64925)
5. (task?switching or task-switching or mental* flexib* or cognitive* flexib* or psychological* flexib* or cognitive* inflexib* or mental* inflexib* or psychological* inflexib* or set?shifting or set?shift*).mp. (5052)
6. 1 or 2 or 3 or 4 or 5 (69322)
7. self report/ (26758)
8. (cognitive flexibility scale? or cognitive flexibility inventor* or attributional style question?aire? or "detail and flexibility question?aire" or DFlex or "coping style? and flexib* inventor*").mp. (140)
9. (self?report* or self report*).mp. (148801)
10. 7 or 8 or 9 (148925)
11. 6 and 10 (816)

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2021.102061.

References

Abbate-Daga, G., Buzzichelli, S., Amianto, F., Rocca, G., Marzola, E., McClintock, S. M., & Fassino, S. (2011). Cognitive flexibility in verbal and nonverbal domains and decision making in anorexia nervosa patients: A pilot study. *BMC Psychiatry, 11* (162), 1–8. https://doi.org/10.1186/1471-244X-11-162.

Alfimova, M. V., Golimbet, V., Lezheiko, T., & Korovaitseva, G. (2017). A potential role of the 5-HTTLPR polymorphism in self-reported executive functioning. *Spanish Journal of Psychology, 20*(e-13), 1–9. https://doi.org/10.1017/sjp.2017.6.

Ashendorf, L., & Mccaffrey, R. J. (2008). Exploring age-related decline on the Wisconsin Card sorting Test. *The Clinical Neuropsychologist, 22*(2), 262–272. https://doi.org/10.1080/13854040701218436.

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154.

Aytug, Z. G., Rua, T., Brazeal, D. V., Almaraz, J. A., & González, C. B. (2018). A socio-cultural approach to multicultural experience: Why interactions matter for creative thinking but exposures don't. *International Journal of Intercultural Relations, 64*, 29–42. https://doi.org/10.1016/j.ijintrel.2018.03.004.

Baker, K. S., Georgiou-Karistianis, N., Lampit, A., Valenzuela, M., Gibson, S. J., & Giummarra, M. J. (2018). Computerised training improves cognitive performance in chronic pain: A participant-blinded randomised active-controlled trial with remote supervision. *Pain, 159*(4), 644–655. https://doi.org/10.1097/j.pain.0000000000001150.

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *Clinical Neuropsychologist, 13* (3), 283–292. https://doi.org/10.1076/clin.13.3.283.1743.

Basso, M. R., Lowery, N., Ghormley, C., & Bornstein, R. A. (2001). Practice effects on the Wisconsin Card sorting Test-64 card version across 12 months. *Clinical Neuropsychologist, 15*(4), 471–478. https://doi.org/10.1076/clin.15.4.471.1883.

Bellinger, D. C., Watson, C. G., Rivkin, M. J., Robertson, R. L., Roberts, A. E., Stopp, C., & Newburger, J. W. (2015). Neuropsychological status and structural brain imaging in adolescents with single ventricle who underwent the fontan procedure. *Journal of the American Heart Association, 4*(12), Article e002302. https://doi.org/10.1161/JAHA.115.002302.

Benton, L. A., Hamsher, K., & Sivan, A. B. (1994). Controlled Oral Word Association Test. *Multilingual Aphasia Examination, 3*.

Ben-Zion, Z., Fine, N. B., Keynan, N. J., Admon, R., Green, N., Halevi, M., & Shalev, A. Y. (2018). Cognitive flexibility predicts PTSD symptoms: Observational and interventional studies. *Frontiers in Psychiatry, 9*, 477. https://doi.org/10.3389/fpsyt.2018.00477.

Berman, N. C., Wheaton, M. G., McGrath, P., & Abramowitz, J. S. (2010). Predicting anxiety: The role of experiential avoidance and anxiety sensitivity. *Journal of Anxiety Disorders, 24*(1), 109–113. https://doi.org/10.1016/j.janxdis.2009.09.005.

Berryman, C., Stanton, T. R., Bowering, K. J., Tabor, A., McFarlane, A., & Moseley, G. L. (2014). Do people with chronic pain have impaired executive function? *A meta-analytical review. Clinical Psychology Review, 34*(7), 563–579. https://doi.org/10.1016/j.cpr.2014.08.003.

Bond, F. W., Hayes, S. C., Baer, R., & A., C., Kenneth (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*(4), 676–688. https://doi.org/10.1016/j.beth.2011.03.007.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). In *Introduction to meta-analysis*. Chichester, United Kingdom: John Wiley & Sons Ltd.

Brenner, L. (2017). *The relationship between measures of cognitive flexibility and intimate partner violence perpetration in a court-mandated domestic violence treatment program* (Doctoral disssertation). Retrieved from ProQuest Dissertations and Theses (10264484).

Buchsbaum, B. R., Greer, S., Chang, W. L., & Berman, K. F. (2005). Meta-analysis of neuroimaging studies of the Wisconsin Card-Sorting task and component processes. *Human Brain Mapping, 25*(1), 35–45. https://doi.org/10.1002/hbm.20128.

Bunk, S., Preis, L., Zuidema, S., Lautenbacher, S., & Kunz, M. (2019). Executive functions and pain. *Journal of Neuropsychology, 30*(3), 169–196. https://doi.org/10.1024/1016-264X/a000264.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton tests*.

Burmester, B., Leathem, J., & Merrick, P. (2016). Subjective cognitive complaints and objective cognitive function in aging: A systematic review and meta-analysis of recent cross-sectional findings. *Neuropsychological Review, 26*(4), 376–393. https://doi.org/10.1007/s11065-016-9332-2.

Canas, J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics, 46*(5), 482–501. https://doi.org/10.1080/0014013031000061640.

Carleton, R. N., Norton, M. A., & Asmundson, G. J. (2007). Fearing the unknown: A short version of the intolerance of uncertainty scale. *Journal of Anxiety Disorders, 21*(1), 105–117. https://doi.org/10.1016/j.janxdis.2006.03.014.

Clayson, K. A. (2018). *A multimethod examination of the relevance of executive control to disgust and mental contamination among female assault survivors* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (106173522018).

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin, 112*(1), 155–159.

Cohen, J. N., Jensen, D., Stange, J. P., Neuburger, M., & Heimberg, R. G. (2017). The immediate and long-term effects of an intensive meditation retreat. *Mindfulness, 8* (4), 1064–1077. https://doi.org/10.1007/s12671-017-0682-5.

Corp, I. B. M. (2011). *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences, 38*(9), 571–578. https://doi.org/10.1016/j.tins.2015.07.003.

Dassen, F. C. M., Houben, K., Allom, V., & Jansen, A. (2018). Self-regulation and obesity: The role of executive function and delay discounting in the prediction of weight loss. *Journal of Behavioral Medicine, 41*(6), 806–818. https://doi.org/10.1007/s10865-018-9940-9.

Davis, J. C., Marra, C. A., Najafzadeh, M., & Liu-Ambrose, T. (2010). The independent contribution of executive functions to health related quality of life in older women. *BMC Geriatrics, 10*(16), 1–8. https://doi.org/10.1186/1471-2318-10-16.

Dawson, E. L., Shear, P. K., & Strakowski, S. M. (2012). Behavior regulation and mood predict social functioning among healthy young adults. *Journal of Clinical and Experimental Neuropsychology, 34*(3), 297–305. https://doi.org/10.1080/13803395.2011.639297.

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior, 31*, 271–327. https://doi.org/10.1016/S0065-2407(03)31007-9.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System*.

Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: An update. *Journal of the International Neuropsychological Society, 10*(2), 301–303. https://doi.org/10.1017/S1355617704102191.

Dennis, J. P., & Vander Wal, J. S. (2010). The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research, 34*(3), 241–253. https://doi.org/10.1007/s10608-009-9276-4.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750.

Dingemans, A. E., Vanhaelen, C. B., Aardoom, J. J., & van Furth, E. F. (2019). The influence of depressive symptoms on executive functioning in binge eating disorder: A comparison of patients and non-obese healthy controls. *Psychiatry Research, 274*, 138–145. https://doi.org/10.1016/j.psychres.2019.02.033.

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open, 6*(12), Article e011458. https://doi.org/10.1136/bmjopen-2016-011458.

Drake, C. E. (2008). *Evaluating a behavioral measure of psychological flexibility (Doctoral dissertation)*. Retrieved from ProQuest Dissertations and Theses (3329528).

Etkin, A., Gyurak, A., & O'Hara, R. (2013). A neurobiological approach to the cognitive deficits of psychiatric disorders. *Dialogues in Clinical Neuroscience, 15*(4), 419–429. https://doi.org/10.31887/DCNS.2013.15.4/aetkin.

Figueroa, I. J., & Youmans, R. J. (2013). Failure to maintain set: A measure of distractibility or cognitive flexibility? *Proceedings of the Human Factors and Ergonomics Society, 57*(1), 828–832. https://doi.org/10.1177/1541931213571180.

Eslinger, P. J., & Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia, 31*(1), 17–28. https://doi.org/10.1016/0028 3932(93)90077-D.

Friederich, H. C., & Herzog, W. (2011). Cognitive-behavioral flexibility in anorexia nervosa. In R. A. H. Adan, & W. H. Kaye (Eds.), *Behavioural Neurobiology of Eating Disorders* (pp. 111–123). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-15131-6.

Fujino, J., Tei, S., Itahashi, T., Aoki, Y., Ohta, H., Kubota, M., … Takahashi, H. (2019). Need for closure and cognitive flexibility in individuals with autism spectrum disorder: A preliminary study. *Psychiatry Research, 271*, 247–252. https://doi.org/10.1016/j.psychres.2018.11.057.

Gajewski, P. D., Ferdinand, N. K., Kray, J., & Falkenstein, M. (2018). Understanding sources of adult age differences in task switching: Evidence from behavioral and ERP studies. *Neuroscience and Biobehavioral Reviews, 92*, 255–275. https://doi.org/10.1016/j.neubiorev.2018.05.029.

Gelonch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and objective measures. *Comprehensive Psychiatry, 66*, 113–122. https://doi.org/10.1016/j.comppsych.2016.01.002.

Gocłowska, M. A., Crisp, R. J., & Labuschagne, K. (2012). Can counter-stereotypes prime flexible thinking? *Group Processes and Intergroup Relations, 16*, 217–231.

Golden, C. J. (1978). *Stroop Color and Word Test: A manual for clinical and experimental uses*. Chicago: Stoelting.

Gonzalez, C. A., Figueroa, I. J., Bellows, B. G., Rhodes, D., & Youmans, R. J. (2013). A new behavioral measure of cognitive flexibility. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 297–306). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39360-0-33.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 38*(4), 404–411. https://doi.org/10.1037/h0059831.

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin, 53*(4), 267–293. https://doi.org/10.1037/h0040755.

Guilford, J. P. (1959). *Personality*. New York:: McGraw-Hill.

Guilford, J. P., & Guilford, J. S. (1980). *Match Problems: Manual of instructions and interpretations*. Sheridan Psychological Services.

Guy, S. C., Isquith, P. K., & Gioia, G. A. (2004). *BRIEF SR: Behavior Rating Inventory of Executive Function self Report Version*. Professional Manual. Psychological Assessment Resources.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and Commitment Therapy: Model, processes and outcomes. *Behaviour Research and Therapy, 44*(1), 1–25. https://doi.org/10.1016/j.brat.2005.06.006.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G., & Curtiss, G. (1993). *Wisconsin Card sorting Test manual: Revised and expanded*. Psychological Assessment Resources.

Hertzberg, L. C. (2017). *The differential roles of objective neuropsychological testing and self report measures in assessing neurocognitive impairment in relation to addiction severity*

among alcohol abusing individuals (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (10287393).
Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (2019). *Cochrane Handbook for Systematic Reviews of Interventions* (6th ed.) Retrieved from www.training.cochrane.org/handbook.
Hoffmann, D., Rask, C. U., & Frostholm, L. (2019). Acceptance and commitment therapy for health anxiety. In *The Clinician's Guide to Treating Health Anxiety* (pp. 123–142). https://doi.org/10.1016/B978-0-12-811806-1.00007-X.
Holliday, J., Tchanturia, K., Landau, S., Collier, D., & Treasure, J. (2005). Is impaired set-shifting an endophenotype of anorexia nervosa? *American Journal of Psychiatry, 162* (12), 2269–2275. https://doi.org/10.1176/appi.ajp.162.12.2269.
Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology, 30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001.
Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2013). The role of cognitive flexibility in cognitive restructuring skill acquisition among older adults. *Journal of Anxiety Disorders, 27*(6), 576–584. https://doi.org/10.1016/j.janxdis.2012.10.004.
Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). Reliability and validity of two self-report measures of cognitive flexibility. *Psychological Assessment, 26*(4), 1381–1387. https://doi.org/10.1037/a0038009.
Kanagaratnam, P., & Asbjørnsen, A. E. (2007). Executive deficits in chronic PTSD related to political violence. *Journal of Anxiety Disorders, 21*(4), 510–525. https://doi.org/10.1016/j.janxdis.2006.06.008.
Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skew and kurtosis. *Restorative Dentistry & Endodontics, 38*(1), 52–54. https://doi.org/10.5395/rde.2017.42.2.152.
Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., & Frank, M. C. (2018). A practical guide for transparency in psychological science. Collabra. *Psychology, 4*(1), 2–15. https://doi.org/10.1525/collabra.158.
Kramer, J. H., Mungas, D., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., & Widmeyer, M. (2014). NIH EXAMINER: Conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society, 20* (1), 11–19. https://doi.org/10.1017/S1355617713001094.
Kuula, L., Pesonen, A. K., Heinonen, K., Kajantie, E., Eriksson, J. G., Andersson, S., & Raikkonen, K. (2018). Naturally occurring circadian rhythm and sleep duration are related to executive functions in early adulthood. *Journal of Sleep Research, 27*(1), 113–119. https://doi.org/10.1111/jsr.12581.
Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials, 2*(2), 93–113. https://doi.org/10.1016/0197-2456(81)90001-5.
Lang, K., Stahl, D., Espie, J., Treasure, J., & Tchanturia, K. (2014). Set shifting in children and adolescents with anorexia nervosa: An exploratory systematic review and meta-analysis. *International Journal of Eating Disorders, 47*(4), 394–399. https://doi.org/10.1002/eat.22235.
Lange, F. (2020). Are difficult-to-study populations too difficult to study in a reliable way: Lessons learned from meta-analyses in clinical neuropsychology. *European Psychologist, 25*(1), 41–50. https://doi.org/10.1027/1016-9040/a000384.
Lange, F., & Dewitte, S. (2019). Cognitive flexibility and pro-environmental behaviour: A multimethod approach. *European Journal of Personality, 33*(4), 488–505. https://doi.org/10.1002/per.2204.
Lau, K. (2012). *The role of cognitive flexibility in acculturation and bicultural stress among Chinese american adults (Doctoral dissertation)*. Retrieved from ProQuest Dissertations and Theses (3517364).
Lee, H., Lamb, S. E., Bagg, M. K., Toomey, E., Cashin, A. G., & Moseley, G. L. (2018). Reproducible and replicable pain research. *Pain, 159*(9), 1683–1689. https://doi.org/10.1097/j.pain.0000000000001254.
Lee, J., & Orsillo, S. (2014). Investigating cognitive flexibility as a potential mechanism of mindfulness in generalized anxiety disorder. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 208–216. https://doi.org/10.1016/j.jbtep.2013.10.008.
Lee, R., Hermens, D., Porter, M., & Redoblado-Hodge, A. (2012). A meta-analysis of cognitive deficits in first-episode major depressive disorder. *Journal of Affective Disorders, 140*(2), 113–124. https://doi.org/10.1016/j.jad.2011.10.023.
Leppanen, J., Adamson, J., & Tchanturia, K. (2018). Impact of cognitive remediation therapy on neurocognitive processing in anorexia nervosa. *Frontiers in Psychiatry, 9* (96), 1–5. https://doi.org/10.3389/fpsyt.2018.00096.
Leung, R. C., & Zakzanis, K. K. (2014). Brief report: Cognitive flexibility in autism spectrum disorders: A quantitative review. *Journal of Autism and Developmental Disorders, 44*(10), 2628–2645. https://doi.org/10.1007/s10803-014-2136-4.
Lounes, N., Khan, G., & Tchanturia, K. (2011). Assessment of cognitive flexibility in anorexia nervosa-self-report or experimental measure? A brief report. *Journal of the International Neuropsychological Society, 17*(5), 925–928. https://doi.org/10.1017/S1355617711000671.
Lüdecke, D. (2019). *esc: Effect size computation for meta analysis (version 0.5.1)*. https://doi.org/10.5281/zenodo.1249218.
Martin, M. M., & Anderson, C. M. (1998). The cognitive flexibility scale: Three validity studies. *International Journal of Phytoremediation, 11*(1), 1–9. https://doi.org/10.1080/08934219809367680.
Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports, 76*(2), 623–626. https://doi.org/10.2466/pr0.1995.76.2.623.
McCloskey, L. (2010). Response frequency and Exner's active-passive ratio as Rorschach indicators of flexibility. *Perceptual and Motor Skills, 111*(2), 559–575. https://doi.org/10.2466/02.03.04.09.PMS.111.5.559-575.
Mednick, S. A. (1968). The Remote Associates Test. *The Journal of Creative Behavior, 2*(3), 213–214.
Meltzer, E. P., Kapoor, A., Fogel, J., Elbulok-Charcape, M. M., Roth, R. M., Katz, M. J., & Rabin, L. A. (2017). Association of psychological, cognitive, and functional variables with self-reported executive functioning in a sample of nondemented community-dwelling older adults. *Applied Neuropsychology: Adult, 24*(4), 364–375. https://doi.org/10.1080/23279095.2016.1185428.
Miles, S., Gnatt, I., Phillipou, A., & Nedeljkovic, M. (2020). Cognitive flexibility in acute anorexia nervosa and after recovery: A systematic review. *Clinical Psychology Review, 81*(101905), 1–129. https://doi.org/10.1016/j.cpr.2020.101905.
Miles, S., Howlett, C. A., Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the Wisconsin Card sorting Test to assess cognitive flexibility. *Behavior Research Methods, 1–9*. https://doi.org/10.3758/s13428-021-01551-3.
Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language, 21*(2), 169–183. https://doi.org/10.1055/s-2000-7563.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., & Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097.
National Heart, Lung, and Blood Insitute. (2014). *Quality assessment tool for observational cohort and cross-sectional studies*. Retrieved from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools.
Nguyen, L., Murphy, K., & Andrews, G. (2019). Cognitive and neural plasticity in old age: A systematic review of evidence from executive functions cognitive training. *Ageing Research Reviews, 53*, Article 100912. https://doi.org/10.1016/j.arr.2019.100912.
Palm, K. M., & Follette, V. M. (2011). The roles of cognitive flexibility and experiential avoidance in explaining psychological distress in survivors of interpersonal victimization. *Journal of Psychopathology and Behavioral Assessment, 33*(1), 79–86. https://doi.org/10.1007/s10862-010-9201-x.
Perpiná, C., Segura, M., & Sánchez-Reales, S. (2017). Cognitive flexibility and decision-making in eating disorders and obesity. *Eating and Weight Disorders, 22*(3), 435–444. https://doi.org/10.1007/s40519-016-0331-3.
Peterson, C., Semmel, A., von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. P. (1982). The attributional style questionnaire. *Cognitive Therapy and Research, 6*(3), 287–299. https://doi.org/10.1007/BF01173577.
R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.
Rabin, L., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA division 40 members. *Archives of Clinical Neuropsychology, 20*(1), 33–65. https://doi.org/10.1016/j.acn.2004.02.005.
Rabin, L., Roth, R., Isquith, P., Wishart, H., Nutterupham, K., Pare, N., & Saykin, A. (2006). Self- and informant reports of executive function on the BRIEF-A in MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology, 21*(7), 721–732. https://doi.org/10.1016/j.acn.2006.08.004.
Reed, R. G., Combs, H. L., & Segerstrom, S. C. (2020). The structure of self-regulation and its psychological and physical health correlates in older adults. *Collabra Psychology, 6* (1), 23. https://doi.org/10.1525/collabra.297.
Reitan, R. M. (1958). Validity of the Trail making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*, 271–276.
Rende, B. (2000). Cognitive flexibility: Theory, assessment, and treatment. *Seminars in speech and language, 21*(2), 121–153. https://doi.org/10.1055/s-2000-7560.
Resnik, D. B. (2007). Conflicts of interest in scientific research related to regulation or litigation. *The Journal of Philosophy, Science & Law, 7*(1), 1–12. https://doi.org/10.5840/jpsl2007722.
Review Manager (RevMan) [Computer program]. (2014). *Version 5.3*. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration.
Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., Mcinnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society, 4*(5), 474–490. https://doi.org/10.1017/S1355617798455073.
Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L., & Treasure, J. (2007). A systematic review and meta-analysis of set-shifting ability in eating disorders. *Psychological Medicine, 37*(8), 1075–1084. https://doi.org/10.1017/S0033291707009877.
Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). Behavior Rating Inventory of Executive Function - Adult Version: Professional manual. *Psychological Assessment Resources*. https://doi.org/10.1109/icgse.2012.22.
Saylik, R., & Szameitat, A. J. (2018). The association between negative attributional style and working memory performance. *The Open Psychology Journal, 11*, 131–141. https://doi.org/10.2174/1874350101811010131.
Schaie, W. K., & Parham, I. A. (1975). *Test of Behavioral Rigidity*. Consulting Psychologists Press.
Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making, 7*(16), 1–6. https://doi.org/10.1186/1472-6947-7-16.
Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., & Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ, 349*, Article g7647. https://doi.org/10.1136/bmj.g7647.
Shepherd, L., Reynolds, D. P., Turner, A., O'Boyle, C. P., & Thompson, A. R. (2019). The role of psychological flexibility in appearance anxiety in people who have experienced a visible burn injury. *Burns, 45*(4), 942–949. https://doi.org/10.1016/j.burns.2018.11.015.

Spencer, R. J., Drag, L. L., Walker, S. J., & Bieliauskas, L. A. (2010). Self-reported cognitive symptoms following mild traumatic brain injury are poorly associated with neuropsychological performance in OIF/OEF veterans. *Journal of Rehabilitation Research and Development, 47*(6), 521–530. https://doi.org/10.1682/JRRD.2009.11.0181.

Steinglass, J., Walsh, B. T., & Stern, Y. (2006). Set shifting deficit in anorexia nervosa. *Journal of the International Neuropsychological Society, 12*(3), 431–435. https://doi.org/10.10170/S1355617706060528.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651.

Tchanturia, K., Anderluh, M. B., Morris, R. G., Rabe-Hesketh, S., Collier, D. A., Sanchez, P., & Treasure, J. L. (2004). Cognitive flexibility in anorexia nervosa and bulimia nervosa. *Journal of the International Neuropsychological Society, 10*(4), 513–520. https://doi.org/10.1017/S1355617704104086.

Tchanturia, K., Campbell, I. C., Morris, R., & Treasure, J. (2005). Neuropsychological studies in anorexia nervosa. *International Journal of Eating Disorders, 37*, S72–S76. https://doi.org/10.1002/eat.20119.

Tchanturia, K., Harrison, A., Davies, H., Roberts, M., & Oldershaw, A. (2011). Cognitive flexibility and clinical severity in eating disorders. *PLoS ONE, 6*(6), 20462. https://doi.org/10.1371/journal.pone.0020462.

Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., & Morris, R. (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card sorting Task. *PLoS ONE, 7*(1), Article e28331. https://doi.org/10.1371/journal.pone.0028331.

Thoma, P., Wiebel, B., & Daum, I. (2007). Response inhibition and cognitive flexibility in schizophrenia with and without comorbid substance use disorder. *Schizophrenia Research, 92*(1–3), 168–180. https://doi.org/10.1016/j.schres.2007.02.004.

Vall, E., & Wade, T. D. (2015). Trail Making Task performance in inpatients with anorexia nervosa and bulimia nervosa. *European Eating Disorders Review, 23*(4), 304–311. https://doi.org/10.1002/erv.2364.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. Retrieved from https://www.jstatsoft.org/v036/i03.

Wecker, N. S., Kramer, J. H., Hallam, B. J., & Delis, D. C. (2005). Mental flexibility: Age effects on switching. *Neuropsychology, 19*(3), 345–352. https://doi.org/10.1037/0894-4105.19.3.345.

Westwood, H., Stahl, D., Mandy, W., & Tchanturia, K. (2016). The set-shifting profiles of anorexia nervosa and autism spectrum disorder using the Wisconsin Card sorting Test: A systematic review and meta-analysis. *Psychological Medicine, 46*(9), 1809–1827. https://doi.org/10.1017/S0033291716000581.

Whiting, D. L., Deane, F. P., Simpson, G. K., McLeod, H. J., & Ciarrochi, J. (2017). Cognitive and psychological flexibility after a traumatic brain injury and the implications for treatment in acceptance-based therapies: A conceptual review. *Neuropsychological Rehabilitation, 27*(2), 263–299. https://doi.org/10.1080/09602011.2015.1062115.

Wildes, J. E., Forbes, E. E., & Marcus, M. D. (2014). Advancing research on cognitive flexibility in eating disorders: The importance of distinguishing attentional set-shifting and reversal learning. *International Journal of Eating Disorders, 47*(3), 227–230. https://doi.org/10.1002/eat.22243.

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2018). Cognitive underpinnings of nationalistic ideology in the context of Brexit. *Proceedings of the National Academy of Sciences of the United States of America, 115*(19), E4532–E4540. https://doi.org/10.1073/pnas.1708960115.

Caitlin A. Howlett BPsych(Hons), is a PhD Candidate within Innovation, Implementation and Clinical Translation (IIMPACT) in Health at the University of South Australia. Her research focuses on cognitive flexibility as a potential indicator of vulnerability in people with persistent pain.

Michael A. Wewege MSc, is a PhD candidate in the School of Health Sciences at the University of New South Wales and the Centre for Pain IMPACT at Neuroscience Research Australia. His research uses network meta-analysis to evaluate the comparative effectiveness of analgesic medicines for adults with low back pain.

Carolyn Berryman MAppSc(Phty)(Hons), MScMed (Pain Mgt), PhD, is a physiotherapist with specialist clinical training in pain science and management. She is also an Australian National Health and Medical Research Council (NHMRC) Early Career Research Fellow within the Faculty of Health and Medical Sciences at The University of Adelaide. Her research focuses on uncovering the neurophysiological mechanisms underpinning pain conditions. https://scholar.google.com.au/citations?user AiSZp8QAAAAJ&hl en

Annika Oldach is studying Cognitive Sciences at Osnabrück University, Germany. For this project, she was a Visiting Scholar at IIMPACT in Health at the University of South Australia.

Elizabeth Jennings M.Ost, is a clinical osteopath with a particular interest in evidence synthesis, women's health and persistent pain, and a Research Assistant with IIMPACT in Health at the University of South Australia.

Emily Moore BPsych(Hons), BPhty(Hons) is a PhD candidate within IIMPACT in Health at the University of South Australia. Her research focuses on evidence synthesis and use of psychoeducational strategies for the treatment of pathological pain syndromes.

Emma Karran BAppSc(Phty)(Hons), Grad Dip (Psych), PhD, is a physiotherapist with specialist expertise in the management of spinal pain. She is also a Research Fellow within IIMPACT in Health at the University of South Australia, where her research focusses on the role of social determinants of pain-related outcomes, and the use of health communication and education in care. https://scholar.google.com.au/citations?user LrIe1YEAAAAJ&hl en

Kimberley Szeto B.Clin Ex Phys, B. Hth Sci(Hons), is an exercise physiologist with a particular interest in evidence synthesis, persistent pain, and chronic disease management. She is also a Research Assistant within IIMPACT in Health at the University of South Australia.

Leander Pronk BPhys, MSc, is a physiotherapist with a particular interest in evidence synthesis, persistent low back pain and sports-related injuries, and a Research Assistant within IIMPACT in Health at the University of South Australia.

Stephanie Miles BA, BPsych(Hons), is a PhD Candidate at Swinburne University of Technology. Her research focuses on the role of cognitive flexibility and related mechanisms in the development, maintenance and treatment of anorexia nervosa.

G. Lorimer Moseley AO DSc, PhD, BPhty(Hons), FAAHMS, FACP, HonFFPMANZCA, HonMAPA is Director of IIMPACT in Health and Professor of Clinical Neurosciences at the University of South Australia. He is a Leadership Investigator with the NHMRC of Australia and leads the not for profit initiative Pain Revolution. https://scholar.google.com.au/citations?user O9a6ieoAAAAJ&hl en

CHAPTER FOUR

Chapter 4 is published as:

Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology*, *36*(5), 347-372. https://doi.org/10.1037/neu0000796

This article was reproduced with permission from the American Psychological Association. A signed statement of contribution for this co-authored manuscript can be found in Appendix F.

Neuropsychology

Back to the Drawing Board—The Relationship Between Self-Report and Neuropsychological Tests of Cognitive Flexibility in Clinical Cohorts: A Systematic Review and Meta-Analysis

Caitlin A. Howlett, Michael A. Wewege, Carolyn Berryman, Annika Oldach, Elizabeth Jennings, Emily Moore, Emma L. Karran, Kimberley Szeto, Leander Pronk, Stephanie Miles, and G. Lorimer Moseley

Online First Publication, April 7, 2022. http://dx.doi.org/10.1037/neu0000796

CITATION

Howlett, C. A., Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., Karran, E. L., Szeto, K., Pronk, L., Miles, S., & Moseley, G. L. (2022, April 7). Back to the Drawing Board—The Relationship Between Self-Report and Neuropsychological Tests of Cognitive Flexibility in Clinical Cohorts: A Systematic Review and Meta-Analysis. *Neuropsychology*. Advance online publication. http://dx.doi.org/10.1037/neu0000796

AMERICAN PSYCHOLOGICAL ASSOCIATION

Neuropsychology

© 2022 American Psychological Association
ISSN: 0894-4105

https://doi.org/10.1037/neu0000796

Back to the Drawing Board—The Relationship Between Self-Report and Neuropsychological Tests of Cognitive Flexibility in Clinical Cohorts: A Systematic Review and Meta-Analysis

Caitlin A. Howlett[1], Michael A. Wewege[2, 3], Carolyn Berryman[1, 4], Annika Oldach[5], Elizabeth Jennings[1], Emily Moore[1], Emma L. Karran[1], Kimberley Szeto[1], Leander Pronk[1], Stephanie Miles[6], and G. Lorimer Moseley[1]

[1] Innovation, Implementation and Clinical Translation (IIMPACT) in Health, University of South Australia

[2] School of Health Sciences, Faculty of Medicine and Health, University of New South Wales

[3] Centre for Pain IMPACT, Neuroscience Research Australia, Sydney, Australia

[4] Brain Stimulation, Imaging and Cognition Research Group, School of Medicine, The University of Adelaide

[5] School of Human Sciences, Osnabrück University

[6] Centre for Mental Health, Swinburne University of Technology



Objective: Cognitive flexibility has been previously described as the ability to adjust cognitive and behavioral strategies in response to changing contextual demands. Cognitive flexibility is typically assessed via self report questionnaires and performance on neuropsychological tests in research and clinical practice.

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Caitlin A. Howlett https://orcid.org/0000 0002 4584 8641
Michael A. Wewege https://orcid.org/0000 0002 3283 2149
Carolyn Berryman https://orcid.org/0000 0002 5316 0847
Annika Oldach https://orcid.org/0000 0003 0523 9072
Emily Moore https://orcid.org/0000 0003 0827 3751
Emma L. Karran https://orcid.org/0000 0002 6088 8289
Kimberley Szeto https://orcid.org/0000 0001 9469 9139
Stephanie Miles https://orcid.org/0000 0001 8190 894X
G. Lorimer Moseley https://orcid.org/0000 0002 3750 4945

Caitlin A. Howlett and Stephanie Miles were recipients of Australian Government Research Training Program (RTP) scholarships. Carolyn Berry man was supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship (ID 1127155). Michael A. Wewege was supported by a Postgraduate Scholarship from the NHMRC, a School of Medical Sciences Top Up Scholarship from the University of New South Wales, and a PhD Top Up Scholarship from Neuroscience Research Australia. G. Lorimer Moseley, Emily Moore, and Emma L. Karran were supported by a NHMRC Leadership Investigator Grant (ID 1178444) to G. Lorimer Moseley. Emma L. Karran was supported by a project grant from Lifetime Support Authority of South Australia. Kimberley Szeto was supported by a University Research Support Grant to G. Lorimer Moseley. These funding bodies had no role in the study design, collection, analysis or interpretation of the data, preparation of the article, or the decision to submit the article for publication.

G. Lorimer Moseley has received support from Reality Health, ConnectHealth U.K., Seqirus, Kaiser Permanente, Workers' Compensation Boards in Australia, Europe and North America, AIA Australia, the International Olympic Committee, Port Adelaide Football Club, and Arsenal Football Club. Professional and scientific bodies have reimbursed him for travel costs related to presentation of research at scientific conferences/symposia. He has received speaker fees for lectures on pain and rehabilitation. He receives book royalties from NOIgroup publications, Dancing Giraffe Press, and OPTP. Carolyn Berryman has received support from Workers' Compensation Boards in Australia. All remaining authors declare no conflicts of interest (personal or financial) with respect to the research, authorship, and/or publication of the article.

We would like to thank the authors of the included studies who responded to enquiries for data and/or provided additional information required for inclusion this review would simply not have been possible without their assistance and generosity. We also wish to thank several authors who responded promptly to assist us in determining the eligibility of their paper(s).

Caitlin A. Howlett played a lead role in data curation, investigation, project administration, resources, supervision, validation, visualization, and writing of original draft and an equal role in conceptualization, formal analysis, methodology, software, and writing of review and editing. Michael A. Wewege played a supporting role in writing of review and editing and an equal role in formal analysis and software. Carolyn Berryman played a supporting role in supervision and an equal role in methodology and writing of review and editing. Annika Oldach played a supporting role in investiga tion, project administration, and writing of review and editing. Elizabeth Jennings played a supporting role in investigation and writing of review and editing. Emily Moore played a supporting role in investigation, project administration, and writing of review and editing. Emma L. Karran played a supporting role in supervision and an equal role in methodology and writing of review and editing. Kimberley Szeto played a supporting role in investi gation, project administration, and writing of review and editing. Leander Pronk played a supporting role in investigation, project administration, and writing of review and editing. Stephanie Miles played a supporting role in investigation, project administration, and writing of review and editing. G. Lorimer Moseley played a lead role in supervision and an equal role in conceptualization, methodology, and writing of review and editing.

The data are available at https://osf.io/mxjfu/.
The experiment materials are available at https://osf.io/mxjfu/.
The preregistered design is available at https://osf.io/mxjfu/.
The preregistered design is available at https://osf.io/mxjfu/.

Correspondence concerning this article should be addressed to G. Lorimer Moseley, Innovation, Implementation and Clinical Translation (IIMPACT) in Health, University of South Australia, GPO Box 2471, Adelaide, SA 5001, Australia. Email: lorimer.moseley@gmail.com

A common assumption among researchers and clinicians is that self report and neuropsychological tests of cognitive flexibility assess the same or similar constructs, but the extent of the relationship between these two assessment approaches in clinical cohorts remains unknown. We undertook a systematic review and meta analysis to determine the relationship between self report and neuropsychological tests of cognitive flexibility in clinical samples. ***Method:*** We searched 10 databases and relevant gray literature (e.g., other databases and pearling) from inception to October 2020 and used the Preferred Reporting Items for Systematic Reviews and Meta Analyses reporting guidelines. Eleven articles including 405 participants satisfied our eligibility criteria. ***Results:*** A multilevel random effects meta analysis revealed no relationship between self report and neuropsychological tests of cognitive flexibility (0.01, 95% CI [0.16 to 0.18]). Individual random effects meta analyses between 12 different tests pairs also found no relationship. ***Conclusion:*** Based on our results, it is clear that the two assessment approaches of cognitive flexibility provide independent information they do not assess the same construct. These findings have important ramifications for future research and clinical practice there is a need to reconsider what constructs self report and neuropsychological tests of "cognitive flexibility" actually assess, and avoid the interchangeable use of these assessments in clinical samples.

Key Points

Question: What is the relationship between self report and neuropsychological tests of cognitive flexibility in clinical cohorts? ***Findings:*** Self report and neuropsychological tests of cognitive flexibility do not relate and therefore almost certainly do not assess similar constructs. ***Importance:*** These two assessment approaches should not be considered interchangeable tools of cognitive flexibility and their discussion and use should reflect this. ***Next Steps:*** Much work is required to fully understand the psychometric properties of most tests that have been purported to assess cognitive flexibility. Empirical studies that are scientifically rigourous, transparent, and of high quality are needed to elucidate the construct of cognitive flexibility and how best to assess it.

Keywords: clinical cohorts, cognitive flexibility, neuropsychological test, self report test, systematic review

Supplemental materials: https://doi.org/10.1037/neu0000796.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Cognitive flexibility has been conceptualized as the ability to effectively modify perspectives and cognitive and behavioral strate gies in response to changing task demands or unexpected conditions in the environment (Cañas et al., 2003; Deák, 2004; Diamond, 2013). Previous literature offers a plethora of conceptual definitions, some of which (e.g., a specific ability such as "set shifting") are narrow and likely to limit our understanding of this latent construct (see Ionescu, 2012 for a review). Thus, determining how best to conceptualize and understand cognitive flexibility has proven to be no easy feat. Some have argued that rather than conceptualizing cognitive flexibility as a specific executive ability (e.g., Diamond, 2013), it should be thought of as a construct that relies on, and emerges from, the efficient, complex, and multifaceted interplay of several cognitive processes (Dajani & Uddin, 2015; Ionescu, 2012). Although recent theoretical models have suggested that investigating cognitive flexibility through an embodied and embedded framework will enhance our understand ing of the construct (Ionescu, 2012, 2017), this is rarely how cognitive flexibility has been and continues to be assessed by many researchers and clinicians interested in this construct.

In order to understand cognitive flexibility and its potential role in the development and maintenance of various clinical conditions, psychometrically sound that is, valid and reliable assessments are critical. An abundance of self report questionnaires such as the Cognitive Flexibility Inventory (CFI; Dennis & Vander Wal, 2010) and the Cognitive Flexibility Scale (CFS; Martin & Anderson, 1998; Martin & Rubin, 1995) have been developed to evaluate one's self reported cognitive flexibility. Alternatively, standardized neuropsy chological assessments such as the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948; Heaton et al., 1993) and Trail Making Test (TMT; Reitan, 1958) have wide acceptance as tests of cognitive flexibility. Other neuropsychological tests of cognitive flexibility include, for example, the CatBat task (Eliava, 1964; Tchanturia et al., 2002), the Haptic Illusion task (Tchanturia et al., 2001; Uznadze, 1966), and the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997), but these tests are less popular. Both self report and neuropsychological tests that purport to assess cogni tive flexibility have been used in clinical and research contexts as a way of targeting and evaluating therapeutic interventions, as well as understanding potential contributors to the development and mainte nance of particular clinical symptoms and/or conditions (Arlt et al., 2016; Brockmeyer et al., 2014; Genders & Tchanturia, 2010; Holliday et al., 2005; Roberts et al., 2010; Tchanturia et al., 2007, 2016).

Various psychopathologies and clinical conditions have been associated with impaired cognitive flexibility. Previous meta analyses have reported small to moderate impairments in cognitive flexibility in people with persistent pain (Berryman et al., 2014; Bunk et al., 2019) including fibromyalgia (Bell et al., 2018), relative to healthy controls. Similar findings have been reported in two meta analyses investigating cognitive flexibility in people with eating disorders (Roberts et al., 2007), particularly anorexia nervosa (Westwood et al., 2016). Roberts et al. (2007) found that people with eating disorders display lower levels of cognitive flexibility

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

than healthy individuals, with varied effect estimates (small: d = 0.36 to large: d = 1.05) on neuropsychological tasks such as the TMT part B (small), the CatBat task (moderate; Eliava, 1964; Tchanturia et al., 2002), as well as the WCST (i.e., perseverative errors) and the Haptic Illusion task (both large; Tchanturia et al., 2001; Uznadze, 1966). These findings are consistent with Westwood et al. (2016) who indicated that adults with anorexia nervosa performed worse than healthy controls on the WCST (i.e., perseverative errors; d = 0.48). In what appears to be corroborative results, a recent systematic review found that people with anorexia nervosa have poorer self reported cognitive flexibility and reduced performance in some domains of cognitive flexibility (i.e., perceptual tasks) but not others (i.e., verbal and switching tasks), although conducting a meta analysis was not possible due to the variability of the tests and outcome variables used to assess cognitive flexibility (Miles et al., 2020). Unaffected siblings of people with anorexia nervosa and bulimia nervosa have demonstrated poorer performance than healthy controls on neuropsychological tests of cognitive flexibility (Holliday et al., 2005; Roberts et al., 2010; Tenconi et al., 2010). Similar findings have been reported in first degree relatives of people with obsessive compulsive disorder (OCD; Chamberlain et al., 2007), raising the possibility that poor cognitive flexibility may be a potential heritable risk factor for these illnesses. Those with major depressive disorder (Lee et al., 2012), schizophrenia (Thoma et al., 2007), or autism spectrum disorder (ASD; Westwood et al., 2016) also tend to perform worse on neuropsychological tests of cognitive flexibility relative to their nonclinical counterparts. Sternheim et al. (2014) found that females with subclinical OCD symptoms performed significantly worse on both self report and neuropsychological tests of cognitive flexibility than females without OCD symptoms, while Lieberman et al. (2016) found that poor cognitive flexibility as indexed by the Verbal and Design Fluency subtests of the Delis Kaplan Executive Function System (D KEFS; Delis et al., 2001, 2004) was a significant mediator of the association between intolerance of uncertainty and defensive responding to safety cues in adults diagnosed with panic disorder. Together, these findings suggest that poor cognitive flexibility could potentially be a transdiagnostic feature of various clinical conditions.

Self report and neuropsychological tests of cognitive flexibility have been used as, or implied to be, interchangeable assessments of the same construct (e.g., Abbate Daga et al., 2014; Fujino et al., 2019; Oguz et al., 2019; Tchanturia et al., 2012). Inherent to such perspectives is the idea that the capacities targeted by self report tests of cognitive flexibility underpin the ability to perform neuropsychological tests of cognitive flexibility (Abbate Daga et al., 2014; Holliday et al., 2005; Tchanturia et al., 2012) and are therefore interchangeable in clinical and research contexts. Notably, a lack of relationship between self report and neuropsychological tests has been reported in the broader executive function literature (Burmester et al., 2016; Rabin et al., 2006; Spencer et al., 2010). One explanation offered for why these two approaches do not relate well is that, on the one hand, neuropsychological tests require a specific repertoire of responses and some tasks only have a limited number of trials (notably decreasing reliability due to the high error variance that is associated with neuropsychological assessments), and on the other hand, self report questionnaires provide insight into an individual's thoughts and behaviors in everyday life (notably increasing ecological validity; Dang et al., 2020). Recent evidence raises the possibility that the same may also be true for assessments of cognitive flexibility. Preliminary findings in clinical cohorts (Gelonch et al., 2016; Johnco et al., 2014a; Lounes et al., 2011) and nonclinical cohorts (Gelonch et al., 2016; Gonzalez et al., 2013; Lounes et al., 2011) have cast doubt over whether self report and neuropsychological tests of cognitive flexibility relate at all. Lounes et al. (2011) found no correlation between the CFS and the Brixton Spatial Anticipation Test in healthy controls (n = 49) and in people with anorexia nervosa (n = 45). Similarly, Gelonch et al. (2016) reported no relationship between the Behavior Rating Inventory of Executive Function Adult (BRIEF A) shift subscale and various outcomes on neuropsychological tests in healthy controls and in people with fibromyalgia. Johnco et al. (2014a) reported small correlations between particular self report and neuropsychological tests of cognitive flexibility in their healthy control sample (n = 53), but no correlations were observed between any self report and neuropsychological tests of cognitive flexibility in their clinical sample of older adults with comorbid anxiety and depression (n = 47). We recently investigated self report and neuropsychological tests of cognitive flexibility across healthy cohorts and did not detect a relationship between them (Howlett et al., 2021), but this may well be because including only healthy cohorts places floor and ceiling effects that conceal a relationship.

That these tests were developed for use in clinical cohorts, which are less likely to demonstrate floor and ceiling effects, point to the critical need to comprehensively evaluate how well the two approaches to testing cognitive flexibility relate in those with clinical conditions characterized by observations of cognitive or behavioral rigidity. Whether or not the approaches assess the same underlying construct remains uncertain. This systematic review and meta analysis aimed to fill the substantial gap in knowledge by determining the relationship between self report and neuropsychological tests of cognitive flexibility in clinical cohorts.

Method

This systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) guidelines (Moher et al., 2009; Shamseer et al., 2015; see Figure 1) and the search strategy was developed according to the Population, Intervention, Comparison, Outcome (PICO) framework (Schardt et al., 2007).

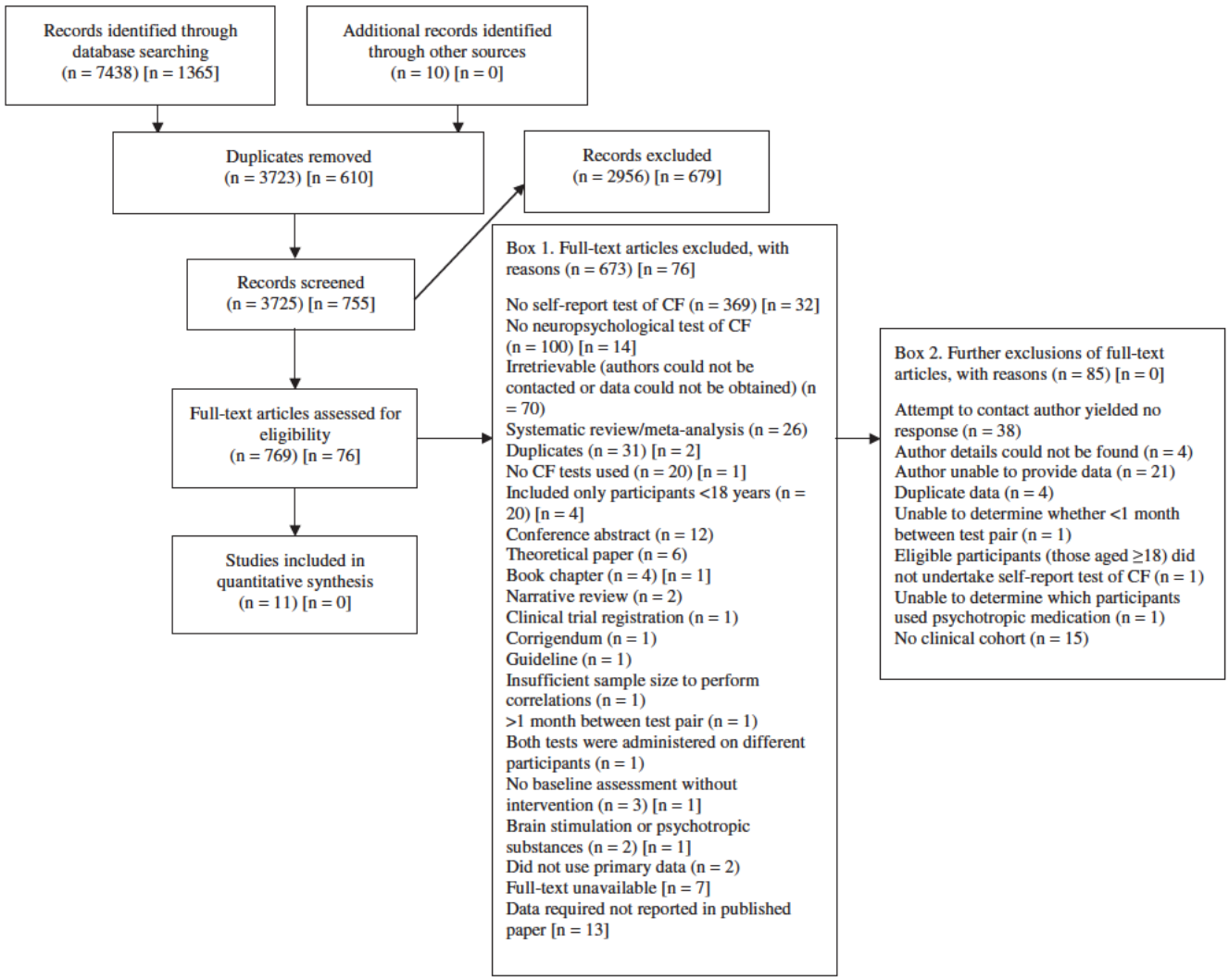
Transparency and Openness

We prospectively registered a protocol on the International Prospective Register of Systematic Reviews; PROSPERO (registration number: CRD42019131567; available from (https://www.crd.york.ac.uk/prospero/display record.php?ID=CRD42019131567) and Open Science Framework (OSF; https://osf.io/mxjfu/); all deviations from our original protocol were recorded, uploaded, and are accessible on OSF. All data, analysis code, and research materials are available on OSF.

Data Sources and Search Strategy

A comprehensive search strategy was developed with assistance from an expert librarian and searches were conducted from inception to 23rd October 2020 in the following databases: MEDLINE (via OvidSP), EMBASE (via OvidSP), EMCARE (via OvidSP), SCOPUS (via Elsevier), PsycINFO (via OvidSP), Web of Science Core

Figure 1
PRISMA Flow Diagram



Note. (*n*) original search; [*n*] updated search; PRISMA Preferred Reporting Items for Systematic Reviews and Meta Analyses.

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Collection (via Clarivate Analytics), Cochrane Central Register of Controlled Trials (CENTRAL), the Cochrane Database of Systematic Reviews, PEDro, and OT seeker. Medical Subject Headings (MeSH) were used for indexing wherever possible and the general search terms included: (cognitive* flexib* OR mental* flexib* OR psychological* flexib* OR wisconsin card? sort* OR trail mak*) AND (cognitive flexibility scale? OR cognitive flexibility inventor* OR self?report*). The search terms were adjusted to suit the capabilities of each database. Gray literature was searched from inception to August 29, 2019. Searches were performed in Trove, ProQuest, the Australian New Zealand Clinical Trials Registry (ANZCTR), Clinicaltrials.gov, the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP), Google Scholar (first 100 references), as well as relevant conference proceedings (see Supplemental File S1 for the full search strategy).

All searches were conducted by a single reviewer (CAH). There were no restrictions on language, country, publication date, nor publication status of the studies. We identified key experts within our field via SCOPUS and contacted these individuals to ensure that any potentially eligible studies that had not been captured by our search were identified. We manually searched the reference lists of all included studies to identify additional studies. The reference lists of all relevant systematic reviews were also hand searched for records containing primary data. For articles with titles that appeared relevant, full text records were retrieved and screened against the predetermined eligibility criteria.

Eligibility Criteria

To be included in this review, studies of any design had to either investigate the relationship between, or have access to data for, at least one self report test of cognitive flexibility (e.g., the CFS) and at least one neuropsychological test of cognitive flexibility (e.g., the WCST). In instances where studies used tests to which we were

naïve, we included these tests if the authors explicitly stated a priori (i.e., in the Introduction or Method section of the article) that the test assessed cognitive flexibility. Studies that used cognitive flexibility as a post hoc interpretation of their findings were excluded from this review. For tests that were considered accepted tests of cognitive flexibility for example, the WCST we decided (a priori) which outcome variables to select on the basis of previous literature; however, for tests to which we were naïve, we discussed the best outcome(s) to select until consensus reached and we sought clarification with the author(s) where necessary.

Studies were excluded based on the following criteria:

- Participants <18 years of age
- Did not use at least one self report test and at least one neuropsychological test designed to assess cognitive flexibility
- Included only healthy participants
- Unable to provide correlational or raw data where it was not reported in the original article
- The time frame in which participants completed the test pair exceeded 1 month
- The test pair was separated by an intervention
- The test pair was administered on different participants
- Participants using psychotropic substances at the time of testing, or intervention studies using psychotropic substances at baseline
- Intervention studies using neural stimulation at baseline
- Postintervention data of studies using psychotropic substances or neural stimulation as part of the main intervention

In addition, commentaries, conference abstracts, editorials, protocols, and letters to the editor were only included if a full text article could be sourced and/or if a full data set were able to be obtained from the author(s).

Study Selection

All citations identified by our search were retrieved and exported to a citation management software, Endnote (Clarivate Analytics, Philadelphia). The records were then imported into Covidence (Veritas Health Innovation Ltd, Australia) where duplicates were removed and the remaining records were screened on the basis of title and abstract by two reviewers (of CAH, LP, EJ, SM, EM, and KS). Irrelevant citations were removed. Full texts of articles potentially fulfilling the inclusion criteria were retrieved and were formally screened by two reviewers (of CAH, LP, EJ, SM, AO, EM, and KS). Disagreements were either resolved by an additional reviewer (CAH, EM, or KS), providing the same reviewer had not previously voted on the inclusion/exclusion of an article, or through discussion (with GLM) if consensus could not be reached. Where entire conference proceedings were retrieved, key words such as "flexib," "wisconsin," "trail mak," and "stroop" were searched within the proceedings. For all abstracts that appeared relevant, we contacted the authors to request a full text and/or available data, and eligibility criteria were clarified where necessary. In instances where data were reported as an overall score of executive function (e.g., BRIEF A), we contacted authors to try to obtain the cognitive flexibility data in isolation (e.g., BRIEF A shift subscale). Similarly, for citations where full texts were unavailable, authors were contacted to clarify eligibility and/or to provide full texts or necessary data (raw or aggregate correlational data).

Risk of Bias Assessment

To assess risk of bias, a customized risk of bias form was developed and piloted a priori by combining relevant items from two existing tools: the Appraisal Tool for Cross Sectional Studies (Downes et al., 2016) and the National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross Sectional Studies (National Heart, Lung, & Blood Insitute, 2014). Study quality was independently assessed by two reviewers [CAH (all articles) and either LP, AO, EM or KS] and assessments were compared for level of agreement. For all included studies, each item on the risk of bias tool was judged as having either high bias, low bias, or unclear bias. Where particular items were not relevant to a study, these items were marked as being not applicable. Disagreements between reviewers were addressed through discussion; however, in instances where consensus could not be adequately reached, a third reviewer (GLM) was consulted.

Data Extraction

Data extraction was performed using a piloted, customized form. Two reviewers [CAH (all studies) and either LP, AO, EM or KS] independently extracted the following data: citation details, study characteristics [i.e., aim(s), study design, country, language in which tests were administered, study conditions, setting, sample size (recruited and analyzed), participant inclusion and exclusion criteria], participants demographics (i.e., sex and/or gender, clinical condition, means and standard deviations for age and years of education), details about the assessment pair (i.e., test names and outcome variables), direction of the test outcome (i.e., higher/lower scores indicate better/worse performance), length of time between administration of each test pair, time point of assessment (i.e., no intervention, preintervention, postintervention), intervention details (if applicable), definition of cognitive flexibility (if reported), inferential statistics (i.e., Pearson's r or Spearman's ρ, p values, confidence intervals, and/or standard errors), descriptive statistics of assessments (i.e., means and standard deviation), and the reporting of sample size calculations. Data extraction forms were compared to ensure accuracy between the two reviewers. Authors of dissertations were contacted prior to the commencement of data extraction and were given one week to respond in order to be included in this review. If further information was required than that contained in the full text article, we contacted the authors a maximum of three times over the course of six weeks. If all three attempts were unsuccessful or the authors were unable to provide the required data for quantitative synthesis, the missing information was deemed irretrievable, and the study was excluded.

Where the raw data were obtained by the authors, Statistical Package for Social Sciences (SPSS) Version 20.0 (IBM Corp, 2011) was used to perform correlational analyses, and normality of the data was determined using the methods described by Kim (2013). In

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

instances where the assumptions of normality were violated, Spearman's ρ coefficients were used rather than Pearson's *r* coefficients. Two reviewers (CAH and KS) independently completed this process and the outputs were compared for accuracy.

Data Analysis

We conducted statistical analyses in R (R Core Team, 2019) and used the *metafor* package (Viechtbauer, 2010) to perform a multilevel random effects meta analysis (https://osf.io/mxjfu/). We deviated from our a priori protocol because RevMan 5.3 software (The Cochrane Collaboration, 2014) was unable to perform the required analyses. Multilevel meta analyses allow the analysis of all eligible effects from included studies, accounting for within study dependencies of outcomes that are undertaken by the same participants at different time points (Del Re, 2015). Level 1 represented within study sampling variance, Level 2 represented variance between effect sizes within a study, and Level 3 represented variance between studies. We fitted the model via restricted maximum likelihood estimation with a compound symmetry variance structure and Knapp and Hartung's adjustment (Assink & Wibbelink, 2016). Correlation coefficients (Pearson's *r* and Spearman's ρ) were converted using a Fisher's *Z* transformation for analysis with the *esc* package (Lüdecke, 2019) and converted back to correlation coefficients for presentation of the forest plots (Borenstein et al., 2009). If the scores for a test pair were calculated in opposite directions such that higher scores indicate poor performance on one test and higher scores indicate better performance on the other test, we reversed the scores of one of the tests by multiplying the correlation coefficient by −1, which resulted in both outcomes of the test pair being positive or negative.

Individual random effects meta analyses were performed when at least two studies had data for the same test pair [e.g., BRIEF A shift subscale and TMT part B (time)]. Interpretation of effect sizes were based on the conventions outlined by Cohen (1992): small (0.1), moderate (0.3), and large (>0.5). Statistical significance was determined at an α level of $p < .05$. Heterogeneity was considered statistically significant on the basis of χ^2 (otherwise known as *Q*) $p < .10$, and substantial on the basis of $I^2 > 50\%$ (Higgins et al., 2019).

Moderator and Sensitivity Analyses

We performed a post hoc moderator analysis by age, using a cutoff point of 60 years, in line with the current understanding that cognitive abilities including cognitive flexibility may begin to decline from 60 years (Ashendorf & McCaffrey, 2008; Gajewski et al., 2018; Nguyen et al., 2019). Additionally, we decided a priori to perform a sensitivity analysis based on disclosure of conflicts of interest because nondisclosure can lead to potential bias and favorable results (Resnik, 2007). We undertook the moderator and sensitivity analyses to investigate whether such confounding factors had any impact on the effect estimates.

Results

Retrieved Studies

The initial search identified 8,813 records. Once duplicates (n = 4,333) were removed, 4,480 articles were screened by title and abstract in accordance with our predetermined inclusion/exclusion criteria; 3,635 records were initially excluded. Eight hundred forty five full text articles were retrieved and screened for eligibility. Of these, 749 were excluded (see Figure 1 for a summary of reasons for exclusion). Once full text screening was complete, we contacted the authors of 96 articles to either (a) seek further information required to determine article eligibility and/or (b) request raw or aggregate data. Of the 96 articles, the authors of which were contacted, we excluded a further 85 articles [see Box 2 for full text exclusions (with reasons) in Figure 1], leaving 11 articles eligible for inclusion. Figure 1 indicates the full process of article inclusion and exclusion, and full details on attempts to clarify eligibility and/or obtain data through author correspondence are provided in the Supplemental Files (see Supplemental File S2).

A total of 11 studies satisfied the inclusion criteria and were included in this review. Raw data were obtained from authors of seven articles (Baker et al., 2018a, 2018b; Bellinger et al., 2015; Dassen et al., 2018; Meltzer et al., 2017; Okuda et al., 2017; Whiting et al., 2015). Alternatively, aggregate data were obtained from the authors of three articles (Dingemans et al., 2019; Johnco et al., 2014a; Marceau et al., 2017) and one article contained the required data (Lounes et al., 2011). Any aspect of the eligibility criteria that remained unclear from the published article was clarified through correspondence with the authors. Six studies (Baker et al., 2018a, 2018b; Bellinger et al., 2015; Dingemans et al., 2019; Johnco et al., 2014a; Okuda et al., 2017) reported that a subset of participants were taking psychotropic substances at the time of testing. Where aggregate data were provided, we requested that these participants be removed from the data set; however, where raw data were provided, we omitted these participants from the data set prior to performing statistical analyses. The same approach was used when studies either (a) included participants who were under 18 years of age or (b) when participants did not complete the test pair(s) within 1 month. Consequently, the data presented in this review deviate from that reported in almost all of the published articles. Full details on attempts to obtain data through correspondence with the authors are supplied in the Supplemental Files (see Supplemental File S2).

Attempts were made to contact 112 authors to obtain full texts. Of these 112 authors, we attempted to contact 108 by either email or ResearchGate (ResearchGate GmbH, Berlin, Germany), and four were contacted via phone. We allowed two weeks for these authors to provide either a published paper, unpublished manuscript, and/or the raw or aggregate data. Of these 112 records, 70 were irretrievable because either the author(s) contact details could not be found, they did not reply, or they were unable to provide the necessary data. For the remaining 42 records, the authors were able to provide sufficient information to determine study eligibility, but none were eligible for inclusion. Any non English publications were translated using Google Translate. Almost all records were excluded after attempting to contact authors to obtain full texts [see Box 1 for full text exclusions (with reasons) in Figure 1]; however, three records were not excluded until after clarifying particular aspects of the eligibility criteria and prior to requesting data [see Box 2 for full text exclusions (with reasons) in Figure 1]. Thus, for three records, the authors were contacted twice (See Supplemental File S3).

Risk of Bias of Included Studies

Overall, the included studies were classified as having moderate to high risk of bias predominantly because of the uncertainty surrounding the methods and/or a lack of transparent reporting.

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

The aims/research questions of three studies (Baker et al., 2018b; Johnco et al., 2014a; Lounes et al., 2011) were aligned with the aims of this systematic review. A few results were most notable. First, post hoc power calculations were conducted using an online calculator (Lachin, 1981) with $\alpha = 0.05$ and $\beta = 0.20$ (80% power) and multiple comparisons were controlled for where necessary. None of the included studies had sample sizes that would be considered large enough to detect adjusted power estimates. Second, all studies insufficiently reported the reliability of the test(s) they used. Third, two studies (Dassen et al., 2018; Meltzer et al., 2017) reported that they did not counterbalance the order of the tests across participants and the remaining nine studies did not report either way. Finally, no study reported registering an a priori protocol. An item based summary of the results for the risk of bias assessment is presented in Table 1.

Study Characteristics

From the 11 included studies, there was a total of 405 participants; 295 were females and 110 were males. The included studies had a sample weighted average age of 42.8 ($SD = 9.4$), ranging from 19 to 81 years. Sample sizes from all included cohorts ranged from 5 to 81 participants (Table 2). Four studies used a cross sectional study design; three used a combined cross sectional and correlational design. Cross sectional studies that did not aim to understand associations between variables were considered purely *cross sectional*; cross sectional studies that aimed to understand associations between variables were considered *cross sectional and correlational*. One study was a randomized active controlled trial; one was a controlled sequential groups design; one was a prospective pilot study; and one was a validation study. The clinical conditions that were the focus of each paper were as follows: chronic back pain ($n = 2$), single ventricle lesion ($n = 1$), obesity ($n = 1$), binge eating disorder ($n = 1$), comorbid anxiety and depression ($n = 1$), anorexia nervosa ($n = 1$), substance use disorder ($n = 1$), mild cognitive impairment ($n = 1$), ASD ($n = 1$), acquired brain injury ($n = 1$). Table 2 provides a summary of the key characteristics of each study.

Summary of Self-Report Tests of Cognitive Flexibility

A total of five different self report tests of cognitive flexibility were used across studies. The shift subscale of the BRIEF A (Guy et al., 2004) was the most frequently used self report test ($n = 7$; Baker et al., 2018a, 2018b; Bellinger et al., 2015; Dassen et al., 2018; Dingemans et al., 2019; Marceau et al., 2017; Meltzer et al., 2017). Three studies (Johnco et al., 2014a; Lounes et al., 2011; Okuda et al., 2017) used the CFS (Martin & Anderson, 1998; Martin & Rubin, 1995) and one study (Johnco et al., 2014a) used the CFI (Dennis & Vander Wal, 2010). Whiting et al. (2015) was the only study to use both the Acceptance and Action Questionnaire II (AAQ II; Bond et al., 2011) and the Acceptance and Action Questionnaire – Acquired Brain Injury (AAQ ABI; Sylvester, 2011; see Figure 2).

Summary of Neuropsychological Tests of Cognitive Flexibility

A total of six different neuropsychological tests of cognitive flexibility were used across studies. The TMT (Reitan, 1958) was the most commonly used neuropsychological test of cognitive flexibility ($n = 9$; Baker et al., 2018a, 2018b; Dassen et al., 2018; Dingemans et al., 2019; Johnco et al., 2014a; Marceau et al., 2017; Meltzer et al., 2017; Okuda et al., 2017; Whiting et al., 2015). Six studies (Baker et al., 2018a, 2018b; Dingemans et al., 2019; Johnco et al., 2014a; Meltzer et al., 2017; Okuda et al., 2017) used the Stroop test (Golden, 1975; Stroop, 1935); four studies (Baker et al., 2018a, 2018b; Dingemans et al., 2019; Johnco et al., 2014a) used the WCST (Grant & Berg, 1948; Heaton et al., 1993); two studies (Johnco et al., 2014a; Meltzer et al., 2017) used the Controlled Oral Word Association Test (COWAT; Benton et al., 1994); two studies (Lounes et al., 2011; Okuda et al., 2017) used the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997); one study (Bellinger et al., 2015) used the Verbal Fluency Task; and one study (Marceau et al., 2017) used the Color Word Interference Task of the D-KEFS battery (Delis et al., 2001, 2004). For studies conducted in non-English speaking countries (Dassen et al., 2018; Dingemans et al., 2019; Okuda et al., 2017), we clarified the language that the tests were administered in. Okuda et al. (2017) reported that Japanese versions were used for some tests but not others, while Dassen et al. (2018) and Dingemans et al. (2019) did not report the language of any administered tests in their papers. We clarified this missing information with the authors; Dassen et al. (2018) used the Dutch version of the BRIEF-A and Dingemans et al. (2019) used the Dutch versions of the BRIEF-A, the Stroop test, the TMT, and the WCST (Figure 3).

Analysis of the Relationship Between Self-Report and Neuropsychological Tests

Multilevel Random-Effects Meta-Analysis: Correlations of All Test Pairs Accounting for Dependency of Effect Sizes (Figure 4)

One hundred sixteen comparisons were obtained across 11 studies. Three comparisons (3%) indicated a significant relationship in the predicted direction; seven comparisons (6%) indicated a significant relationship in the direction opposite to that predicted; 106 comparisons (91%) were nonsignificant. The pooled effect estimate was 0.01, 95% CI [−0.16 to 0.18], demonstrating that there was no evidence to suggest that a relationship exists between self report and neuropsychological tests of cognitive flexibility in clinical cohorts (see Figure 4). Our chosen statistical program (R; R Core Team, 2019) was unable to estimate the sampling variance when $n \leq 4$, thus correlations were incomputable in those cases. One study (Baker et al., 2018b) provided us with the raw data whereby only four participants were eligible for inclusion for a particular test pair [i.e., the BRIEF-A shift subscale and the WCST (perseverative errors)], so this test pair was not analyzed (Table 3).

Level of Variance Within the Multilevel Random-Effects Model

Using a log likelihood ratio test, we estimated the variance being contributed to each level of the model and found that when we compared the full model to a reduced model (i.e., a two level model with within study variance removed), the within study variance was nonsignificant ($p = .3$). Thus, we found no significant variability

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Table 1
Risk of Bis Assessment

Study	Were the aim(s) and/or research question(s) stated, and aligned with our research question?	Were the sample demographics (i.e., age, gender/sex) described/obtained?	Was the sample size justified for correlational analyses?	Do the authors report the tests as reliable?	Were the study conditions implemented consistently across all participants?	Were the descriptive data of the tests described/obtained?	Were the results of the correlation described/obtained?	Were planned correlational analyses performed and stated a priori?	Were the relevant results internally consistent?	If parametric tests were used, were the assumptions of normality met for the correlation?	Was the test order counterbalanced across participants and planned for in the analyses?	Were funding sources declared?	Was a conflict of interest statement provided?	Was a protocol registered a priori?
Baker et al. (2018a)	N	Y	N	N	?	Y	Y	N/A	Y	Y	?	Y	Y	N
Baker et al. (2018b)	Y	Y	N	N	?	Y	Y	N/A	Y	Y	?	Y	Y	N
Bellinger et al. (2015)	N	Y	N	N	?	Y	Y	N/A	Y	Y	?	Y	Y	N
Dassen et al. (2018)	N	Y	N	N	N	Y	Y	N/A	Y	N/A	N	Y	Y	N
Dingemans et al. (2019)	N	Y	N	?	?	Y	Y	N/A	Y	?	?	N	Y	N
Johnco et al. (2014a)	Y	Y	N	?	?	Y	Y	?	Y	Y	?	N	Y	N
Lounes et al. (2011)	Y	Y	N	?	?	Y	Y	?	Y	Y	?	Y	Y	N
Marceau et al. (2017)	N	Y	N	N	N	Y	Y	N/A	N	N/A	?	Y	Y	N
Metzer et al. (2017)	N	Y	N	?	Y	Y	Y	N/A	Y	Y	N	Y	Y	N
Okuda et al. (2017)	N	Y	N	?	?	Y	Y	N/A	Y	Y	?	Y	Y	N
Whiting et al. (2015)	N	Y	N	?	?	Y	Y	N/A	Y	N/A	?	N	N	N

Note Y = yes; N = no; ? = unclear; N/A = not applicable; Green = low risk of bias; Red = high risk of bias; Yellow = unclear risk of bias; Blue = not applicable See the online article for the color version of this table

between effect sizes within studies. A subsequent log likelihood ratio test was performed by comparing the full model to a reduced model (i.e., a two level model with between study variance removed) and found that between study variance was significant ($p < .0001$). Thus, we found significant variability between studies. We then calculated the distribution of the total variance across the three levels of the random effects model and found that 36% of the variance is explained by within study sampling variance, 4.4% is between effect sizes within each study, and 59.6% is between study variance.

Moderator Analysis

Dichotomization of "older" and "younger" cohorts were deter mined by whether the mean age of each study sample was above or below 60 years. Two studies comprised the "older" cohort and nine studies comprised the "younger" cohort. Our moderator analysis found no difference in the effects by age group (0.03, 95% CI [−0.45 to 0.51]).

Sensitivity Analysis

Our sensitivity analysis, which considered studies without a conflict of interest statement, revealed no effect (0.02, 95% CI [−0.13 to 0.18]).

Random-Effects Meta-Analysis Between Comparisons of the Same Test Pair(s) From Different Cohorts

Out of 12 possible test pair combinations, none revealed a significant relationship. Forest plots between each of the individual test pairs can be found in the Supplemental Files (see Supplemental File S4).

Discussion

Summary of Findings

This systematic review aimed to determine the relationship between self report and neuropsychological tests of cognitive flex ibility in clinical cohorts. We found no evidence for a relationship, suggesting that self report questionnaires and neuropsychological tests of cognitive flexibility should not be considered as interchange able proxies for the assessment of clinical groups, and researchers and clinicians should be careful not to generalize performance from one assessment to another. This result is based on a multilevel random effects meta analysis and individual random effects meta analyses concerning 12 different test pairs, both of which found no relationship between self report and neuropsychological tests that aimed to assess cognitive flexibility.

Our findings contrast with common understandings but are consistent with previous studies. Johnco et al. (2014a), Gelonch et al. (2016), and Lounes et al. (2011) did not identify a relationship between various self report and neuropsychological tests thought to target cognitive flexibility in people with comorbid anxiety and depression, fibromyalgia, and anorexia nervosa, respectively. Only three studies (Baker et al., 2018b; Johnco et al., 2014a; Lounes et al., 2011) aimed to answer the same research question as ours, and eight studies collected the data to ask this question but did not set out to answer it. That there are not more direct investigations of this matter,

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Table 2
Study Characteristics of Included Studies

Citation and year	Country of origin	Study design	Diagnosis/Condition	*N*	M_{age} (*SD*)	Gender or sex	Mean years of education (*SD*)
Baker et al (2018a)	Australia	Randomized active controlled trial	Chronic back pain (nontraining group)	11	45 0 (16 3)	2 M; 9 F	15 8 (3 5)
			Chronic back pain (training group)	8	42 3 (12 2)	4 M; 4 F	16 6 (5 2)
Baker et al (2018b)	Australia	Cross-sectional	Chronic back pain	5	32 4 (6 1)	1 M; 4 F	13 8 (2 5)
Bellinger et al (2015)	United States of America	Cross-sectional	Single ventricle lesion	15	19 0 (0 6)	10 M; 5 F	—
Dassen et al (2018)	The Netherlands	Cross-sectional, correlational	Obesity	81	41 2 (12 7)	21 M; 60 F	—
Dingemans et al (2019)	The Netherlands	Cross-sectional	Binge eating disorder	55	33 1 (9 7)	3 M; 52 F	—
Johnco et al (2014a)	Australia	Cross-sectional, correlational	Older adults with comorbid anxiety and depression	29	66 8 (4 8)	13 M; 16 F	14 2 (3 5)
Lounes et al (2011)	The United Kingdom	Cross-sectional, correlational	Anorexia nervosa	45	27 6 (7 9)	45 F	—
Marceau et al (2017)	Australia	Controlled sequential groups design	Substance use disorder (CR group)	23 [16]	32 7 (8 8)	23 F	10 3 (2 7)
			Substance use disorder (TAU group)	27 [17]	32 4 (7 5)	27 F	11 2 (2 2)
Meltzer et al (2017)	United States of America	Cross-sectional	Mild cognitive impairment	51	80 8 (6 2)	11 M; 40 F	12 3 (3 5)
Okuda et al (2017)	Japan	Prospective pilot study	Autism spectrum disorder	5	24 0 (1 9)	5 M	14 6 (1 7)
Whiting et al (2015)	Australia	Validation study	Acquired brain injury	50	36 8 (14 1)	40 M; 10 F	11 3 (1 9)

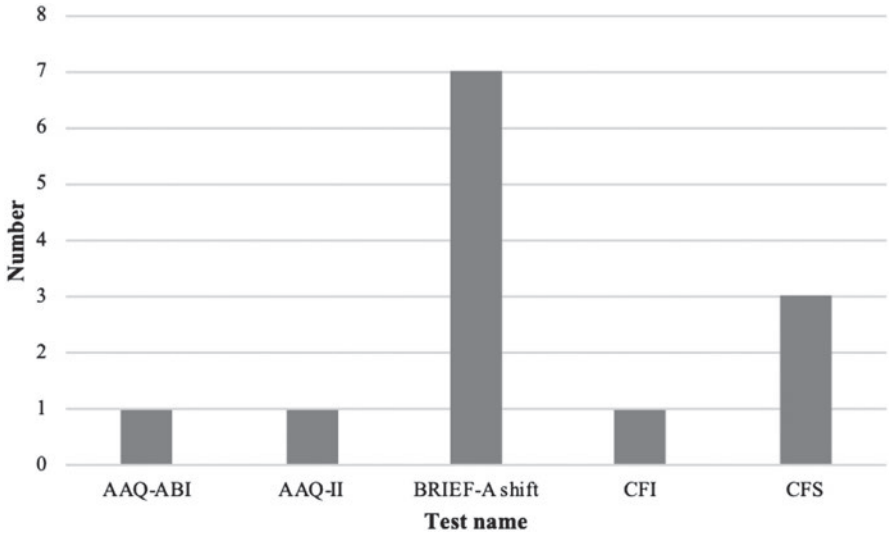
Note. *N* = sample size; *SD* = standard deviation; M = male; F = female; — = information not reported or could not be obtained; M_{age} = mean age; CR = cognitive remediation; TAU = treatment as usual

is in itself is informative because despite common implications of equivalence by the interchangeable use of assessments, the question seems to have been of remarkably little interest to the field. This is particularly interesting when one considers that this assumption of equivalence appears to not have been tested in the very clinical conditions most often considered to be associated with reduced cognitive flexibility traumatic brain injury (Pang et al., 2016) and those with ASD (Fujino et al., 2019; Westwood et al., 2016). Understanding cognitive flexibility has proven challenging even when nonoverlapping aspects of cognitive flexibility have been assessed, particularly when trying to understand its role in condi tions such as ASD (Geurts et al., 2009). However, it is possible that previous studies investigating these clinical groups have relied on one assessment approach to assess cognitive flexibility, or studies that did include both assessments approaches were excluded for reasons beyond our control. Although the relation (or lack thereof) between self report and neuropsychological tests for other executive functions has, more broadly, been of substantial interest to the field (e.g., Dang et al., 2020; Rabin et al., 2006; Spencer et al., 2010; Toplak et al., 2013), the association between self report and neuro psychological tests of cognitive flexibility, more specifically, seems to have received little attention. We recently investigated this matter in healthy adults and found no evidence of a relationship (Howlett et al., 2021), but healthy cohorts are not for whom these tests were typically developed or for whom they guide clinical decisions. The current work solidifies that self report and neuropsychological tests of "cognitive flexibility" are unlikely to assess the same underlying construct.

One reason for the lack of relationship observed in this study could be that the gap between self report and neuropsychological tests of cognitive flexibility is simply too large, particularly when such assessments are often developed with different facets of cognitive flexibility in mind. Moreover, that there were very few test pairs with multiple comparisons may also be due to different accounts of cognitive flexibility being used for various research agendas. This also becomes evident when we consider that particu lar items of self report questionnaires do not appear to directly map onto outcomes of neuropsychological tests that are designed to assess set shifting one component of cognitive flexibility (Tchanturia et al., 2004). For example, the CFS item "I am willing to work at creative solutions to problems" is highly unlikely to tap into identical processes involved in shifting when completing the WCST. Also, self report questionnaires likely draw upon inherently different processes than neuropsychological tests such as meta cognitive awareness and the ability to be perceptive and self reflective. Another potential explanation as to why self report and neuropsychological assessments of cognitive flexibility do not relate well may be partly due to a key shortcoming of neuro psychological tests the *task impurity problem* that is, task per formance relies on a complex network of executive and nonexecutive functions (Buchsbaum et al., 2005; Miyake et al., 2000; Wilde et al., 2014). That neuropsychological assessments do not exclusively assess cognitive flexibility, but rather multiple processes, ultimately makes interpreting and drawing conclusions about task performance exceptionally difficult (Miller & Kerns, 2015). Although Ionescu (2012) has previously made the compel ling argument that "there is no static entity called cognitive flexi bility" (p. 197), attempting to tease apart cognitive flexibility could theoretically compromise the integrity of the construct since its

Figure 2

The Number of Included Studies That Used Each Self Report Test of Cognitive Flexibility



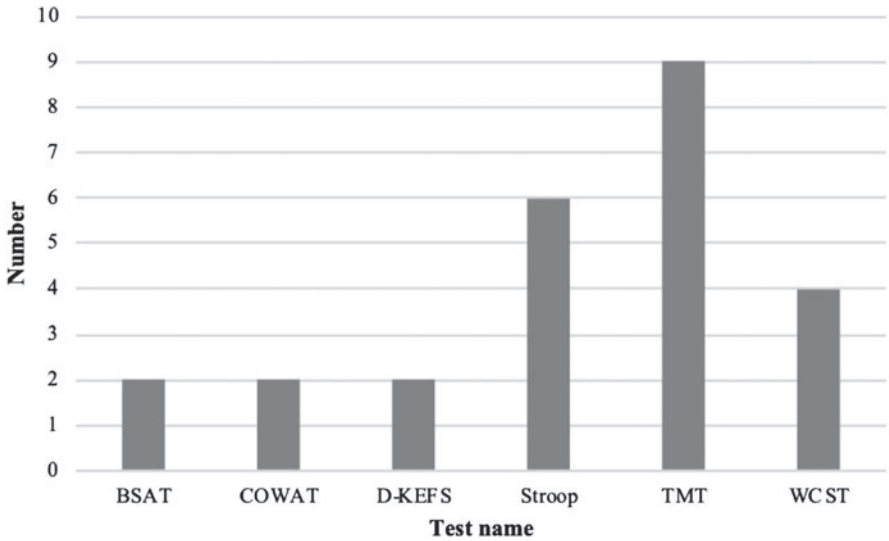
Note. AAQ ABI Acceptance and Action Questionnaire Acquired Brain Injury; AAQ II Acceptance and Action Questionnaire II; BRIEF A shift Behavior Rating Inventory of Executive Function Adult (shift subscale); CFI Cognitive Flexibility Inventory; CFS Cognitive Flexibility Scale.

existence is considered to be an emergent property that fundamentally relies on other cognitive processes (Dajani & Uddin, 2015; Ionescu, 2012).

The primary result is compelling when we look more closely at the data from two studies. For the multilevel and individual random effects meta analysis, we excluded postintervention data from studies with interventions that might be expected to have unclear or differential effects on the two types of tests, for example, psychotropic medications or neurostimulation. However, postintervention data of studies involving cognitive behavioral interventions

Figure 3

The Number of Included Studies That Used Each Neuropsychological Test of Cognitive Flexibility



Note. BSAT Brixton Spatial Anticipation Test; COWAT Controlled Oral Word Association Test; D KEFS Delis Kaplan Executive Function System; Stroop Stroop test; TMT Trail Making Test; WCST Wisconsin Card Sorting Test.

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Table 3

Assessment Details and Correlational Data of Included Studies

Citation and year	Language test(s) were administered in	Self-report test	*M* (*SD*)	Neuropsychological test	*M* (*SD*)	*N* in correlation	Raw correlation	95% CI	*p* value
Baker et al (2018a)[a] Preintervention (noncognitive training group)									
		BRIEF-A/shift subscale[b]	10 8 (3 0)	TMT/Part B time	66 8 (19 2)	11	$r = 291$	95% CI [− 591, 924]	$p = 385$
		BRIEF-A/shift subscale[b]		TMT/Part B-A time[b,c]	34 6 (14 2)	11	$r = 177$	95% CI [− 663, 829]	$p = 603$
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 1 (0 4)	11	$r = -093$	95% CI [− 585, 542]	$p = 786$
		BRIEF-A/shift subscale[b]		WCST/Perseverative errors	18 2 (15 0)	10	$r = -384$	95% CI [− 980, 564]	$p = 273$
		BRIEF-A/shift subscale[b]		Stroop/Color–Word Trial	37 2 (11 5)	11	$r = -335$	95% CI [− 793, 315]	$p = 314$
Postintervention (noncognitive training group)									
		BRIEF-A/shift subscale[b]	10 5 (3 9)	TMT/Part B time	65 0 (17 3)	11	$r = 099$	95% CI [− 676, 669]	$p = 772$
		BRIEF-A/shift subscale[b]		TMT/Part B-A time[b,c]	36 2 (12 4)	11	$r = 065$	95% CI [− 710, 690]	$p = 849$
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 3 (0 4)	11	$r = -177$	95% CI [− 737, 458]	$p = 602$
		BRIEF-A/shift subscale[b]		WCST/Perseverative errors	13 7 (18 8)	10	$\rho = -113$	95% CI [− 748, 701]	$p = 756$
		BRIEF-A/shift subscale[b]		Stroop/Color–Word Trial	39 7 (13 4)	11	$r = -312$	95% CI [− 783, 379]	$p = 350$
Preintervention (cognitive training group)									
		BRIEF-A/shift subscale[b]	12 1 (2 5)	TMT/Part B time	62 3 (22 8)	8	$r = 359$	95% CI [− 491, 984]	$p = 383$
		BRIEF-A/shift subscale[b]		TMT/Part B-A time[b,c]	34 1 (19 7)	8	$r = 391$	95% CI [− 504, 978]	$p = 338$
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 2 (0 7)	8	$r = 380$	95% CI [− 497, 959]	$p = 354$
		BRIEF-A/shift subscale[b]		WCST/Perseverative errors	11 6 (7 4)	7	$r = 069$	95% CI [− 847, 847]	$p = 883$
		BRIEF-A/shift subscale[b]		Stroop/Color–Word Trial	39 3 (12 7)	8	$r = -152$	95% CI [− 913, 728]	$p = 720$
Postintervention (cognitive training group)									
		BRIEF-A/shift subscale[b]	11 8 (3 2)	TMT/Part B time	48 4 (14 6)	8	$r = 304$	95% CI [− 500, 907]	$p = 465$
		BRIEF-A/shift subscale[b]		TMT/Part B-A time[b,c]	25 0 (6 9)	8	$r = 771$	95% CI [455, 987]	$p = 025$
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 2 (0 6)	8	$r = 668$	95% CI [− 366, 958]	$p = 070$
		BRIEF-A/shift subscale[b]		WCST/Perseverative errors	10 4 (8 2)	8	$\rho = 122$	95% CI [− 774, 919]	$p = 774$
		BRIEF-A/shift subscale[b]		Stroop/Color–Word Trial	42 4 (11 4)	8	$r = -194$	95% CI [− 916, 785]	$p = 645$
Baker et al (2018b)[a]		BRIEF-A/shift subscale[b]	10 8 (3 4)	TMT/Part B time	65 4 (28 6)	5	$r = 820$	95% CI [496, 1 00]	$p = 089$
		BRIEF-A/shift subscale[b]		TMT/Part B-A time[b,c]	38 2 (25 0)	5	$\rho = 616$	95% CI [− 667, 1 00]	$p = 269$
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 5 (1 0)	5	$r = 430$	95% CI [−1 00, 1 00]	$p = 470$
		BRIEF-A/shift subscale[b]		WCST/Perseverative errors	11 8 (11 1)	4	$r = 892$	95% CI [841, 1 00]	$p = 108$
		BRIEF-A/shift subscale[b]		Stroop/Color–Word Trial	38 6 (12 6)	5	$\rho = -564$	95% CI [−1 00, 875]	$p = 322$
Bellinger et al (2015)[a]		BRIEF-A/shift subscale[b]	45 0 (9 4)	D-KEFS Verbal Fluency/ Switching accuracy[b]	10 9 (3 5)	15	$r = -312$	95% CI [− 671, 147]	$p = 258$
		BRIEF-A/shift subscale[b]		D-KEFS Verbal Fluency/ Switching versus category[b]	9 5 (3 3)	15	$r = -121$	95% CI [− 605, 327]	$p = 668$
Dassen et al (2018)[a]	BRIEF-A; Dutch version	BRIEF-A/shift subscale	10 5 (2 7)	TMT/Part B time[b,c]	48 9 (21 5)	81	$\rho = 127$	95% CI [− 132, 350]	$p = 258$
		BRIEF-A/shift subscale		TMT/Part B-A time	27 8 (18 1)	81	$\rho = 178$	95% CI [− 074, 396]	$p = 112$
		BRIEF-A/shift subscale		TMT/Part B/A time[b,c]	2 4 (0 8)	81	$\rho = 204$	95% CI [− 030, 429]	$p = 067$

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Table 3 (*continued*)

Citation and year	Language test(s) were administered in	Self-report test	*M* (*SD*)	Neuropsychological test	*M* (*SD*)	*N* in correlation	Raw correlation	95% CI	*p* value
Dingemans et al (2019)[a]	BRIEF-A, Stroop, TMT, WCST; Dutch versions	BRIEF-A/shift subscale[b]	12 2 (3 0)	TMT/Part B time[b,c]	96 4 (30 3)	55	ρ = 045	95% CI [− 222, 306]	*p* = 745
		BRIEF-A/shift subscale[b]		TMT/Part B-A time[b,c]	29 1 (27 3)	56	*r* = − 055	95% CI [− 315, 21]	*p* = 688
		BRIEF-A/shift subscale[b]		TMT/Part B/A time	1 5 (0 5)	56	ρ = − 133	95% CI [− 382, 134]	*p* = 327
		BRIEF-A/shift subscale[b]		WCST/Perseverative errors[c]	18 1 (7 8)	56	ρ = 172	95% CI [− 095, 416]	*p* = 204
		BRIEF-A/shift subscale[b]		WCST/Perseverative responses[b,c]	42 7 (10 4)	55	ρ = 071	95% CI [− 198, 33]	*p* = 609
		BRIEF-A/shift subscale[b]		Stroop/Interference Score/ trial 3 − trial 1[b,c]	32 7 (15 5)	56	ρ = 055	95% CI [− 21, 313]	*p* = 689
		BRIEF-A/shift subscale[b]		Stroop/Trial 3/trial 1	1 6 (0 3)	56	ρ = 015	95% CI [− 248, 276]	*p* = 913
Johnco et al (2014a)[a]		CFI/total score	93 4 (13 9)	TMT/Part B time	88 9 (26 3)	29	*r* = 076	95% CI [− 219, 411]	*p* = 694
		CFI/total score		TMT/Part B-A time[b,c]	57 6 (24 2)	29	*r* = − 037	95% CI [− 309, 335]	*p* = 848
		CFI/total score		TMT/Part B/A time[b,c]	3 0 (0 8)	29	*r* = − 153	95% CI [− 440, 220]	*p* = 427
		CFI/total score		Stroop/Color–Word Trial	33 2 (8 6)	29	*r* = 124	95% CI [− 275, 444]	*p* = 522
		CFI/total score		Stroop/Interference Score[b,c]	−29 1 (9 3)	29	*r* = 043	95% CI [− 321, 357]	*p* = 824
		CFI/total score		COWAT/Total score	37 7 (10 7)	29	*r* = 189	95% CI [− 145, 525]	*p* = 325
		CFI/total score		WCST/Perseverative errors	19 1 (12 1)	29	*r* = − 323	95% CI [− 641, 162]	*p* = 087
		CFI/total score		WCST/% perseverative responses[b]	17 7 (10 8)	29	*r* = − 326	95% CI [− 633, 171]	*p* = 085
		CFI/alternatives subscale	63 6 (11 6)	TMT/Part B time		29	*r* = − 054	95% CI [− 399, 345]	*p* = 781
		CFI/alternatives subscale		TMT/Part B-A time[b,c]		29	*r* = − 160	95% CI [− 467, 285]	*p* = 408
		CFI/alternatives subscale		TMT/Part B/A time[b,c]		29	*r* = − 241	95% CI [− 513, 126]	*p* = 209
		CFI/alternatives subscale		Stroop/Color–Word Trial		29	*r* = 163	95% CI [− 283, 523]	*p* = 398
		CFI/alternatives subscale		Stroop/Interference Score[b,c]		29	*r* = 246	95% CI [− 092, 543]	*p* = 198
		CFI/alternatives subscale		COWAT/Total score		29	*r* = 103	95% CI [− 232, 462]	*p* = 593
		CFI/alternatives subscale		WCST/Perseverative errors		29	*r* = − 240	95% CI [− 593, 242]	*p* = 210
		CFI/alternatives subscale		WCST/% Perseverative responses[b]		29	*r* = − 254	95% CI [− 599, 229]	*p* = 183
		CFI/control subscale	29 8 (6 3)	TMT/Part B time		29	*r* = 267	95% CI [− 087, 566]	*p* = 162
		CFI/control subscale		TMT/Part B-A time[b,c]		29	*r* = 210	95% CI [− 214, 532]	*p* = 274
		CFI/control subscale		TMT/Part B/A time[b,c]		29	*r* = 103	95% CI [− 349, 511]	*p* = 596
		CFI/control subscale		Stroop/Color–Word Trial		29	*r* = − 026	95% CI [− 342, 323]	*p* = 895
		CFI/control subscale		Stroop/Interference Score[b,c]		29	*r* = − 356	95% CI [− 673, 008]	*p* = 058
		CFI/control subscale		COWAT/Total score		29	*r* = 227	95% CI [− 209, 569]	*p* = 236
		CFI/control subscale		WCST/Perseverative errors		29	*r* = − 272	95% CI [− 586, 128]	*p* = 153
		CFI/control subscale		WCST/% Perseverative responses[b]		29	*r* = − 252	95% CI [− 560, 135]	*p* = 188
		CFS/total score	49 3 (8 2)	TMT/Part B time		29	*r* = 033	95% CI [− 314, 461]	*p* = 867
		CFS/total score		TMT/Part B-A time[b,c]		29	*r* = − 072	95% CI [− 438, 368]	*p* = 711
		CFS/total score		TMT/Part B/A time[b,c]		29	*r* = − 232	95% CI [− 592, 193]	*p* = 225
		CFS/total score		Stroop/Color–Word Trial		29	*r* = 238	95% CI [− 177, 564]	*p* = 213
		CFS/total score		Stroop/Interference Score[b,c]		29	*r* = 022	95% CI [− 321, 330]	*p* = 910
		CFS/total score		COWAT/Total score		29	*r* = 184	95% CI [− 194, 560]	*p* = 339
		CFS/total score		WCST/Perseverative errors		29	*r* = − 278	95% CI [− 585, 137]	*p* = 145
		CFS/total score		WCST/% Perseverative responses[b]		29	*r* = − 269	95% CI [− 550, 142]	*p* = 158
Lounes et al (2011)		CFS/total score	45 4 (9 6)	The Brixton Spatial Anticipation Test/Total errors	13 1 (4 9)	45	*r* = 06	95% CI [− 237, 347]	*p* > 05

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Table 3 (*continued*)

Citation and year	Language test(s) were administered in	Self-report test	*M* (*SD*)	Neuropsychological test	*M* (*SD*)	*N* in correlation	Raw correlation	95% CI	*p* value
Marceau et al (2017)[a]									
CR group (preintervention)									
		BRIEF-A/shift subscale[b]	58 3 (10 4)	TMT/Part B time[b,c]	64 4 (20 7)	22	ρ = 069	95% CI [− 347, 498]	*p* = 759
		BRIEF-A/shift subscale[b]		TMT/Part B-A time	38 6 (20 4)	22	ρ = 222	95% CI [− 277, 663]	*p* = 320
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 7 (1 0)	22	ρ = 242	95% CI [− 283, 657]	*p* = 278
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition versus color naming (condition 3 minus condition 1, scaled scores)	9 1 (3 3)	22	ρ = 201	95% CI [− 268, 636]	*p* = 369
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition/Shifting versus color naming (condition 4 minus condition 1, scaled scores)	9 3 (2 9)	22	ρ = 132	95% CI [− 375, 636]	*p* = 558
CR group (postintervention)									
		BRIEF-A/shift subscale[b]	52 1 (9 9)	TMT/Part B time[b,c]	55 6 (20 8)	16	ρ = − 315	95% CI [− 772, 331]	*p* = 235
		BRIEF-A/shift subscale[b]		TMT/Part B-A time	34 7 (20 2)	16	ρ = − 307	95% CI [− 730, 362]	*p* = 248
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 7 (1 1)	16	ρ = − 368	95% CI [− 754, 205]	*p* = 161
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition versus color naming (condition 3 minus condition 1, scaled scores)	11 2 (2 0)	16	ρ = 443	95% CI [− 004, 799]	*p* = 086
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition/shifting versus color naming (condition 4 minus condition 1, scaled scores)	11 4 (2 2)	16	ρ = 314	95% CI [− 224, 719]	*p* = 237
TAU (control) group (preintervention)									
		BRIEF-A/shift subscale[b]	63 8 (14 2)	TMT/Part B time[b,c]	60 9 (22 8)	27	ρ = 157	95% CI [− 290, 484]	*p* = 435
		BRIEF-A/shift subscale[b]		TMT/Part B-A time	34 3 (19 6)	27	ρ = − 127	95% CI [− 485, 252]	*p* = 527
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 4 (0 7)	27	ρ = − 512	95% CI [− 766, − 126]	*p* = 006
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition versus color naming (condition 3 minus condition 1, scaled scores)	10 4 (1 9)	27	ρ = − 189	95% CI [− 570, 262]	*p* = 346
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition/Shifting versus color naming (condition 4 minus condition 1, scaled scores)	10 6 (2 1)	27	ρ = − 363	95% CI [− 740, 083]	*p* = 063
TAU (control) group (postintervention)									
		BRIEF-A/shift subscale[b]	62 1 (9 0)	TMT/Part B time[b,c]	42 7 (5 3)	17	ρ = 253	95% CI [− 248, 621]	*p* = 328
		BRIEF-A/shift subscale[b]		TMT/Part B-A time	21 0 (7 5)	17	ρ = 120	95% CI [− 428, 596]	*p* = 648
		BRIEF-A/shift subscale[b]		TMT/Part B/A time[b,c]	2 1 (0 5)	17	ρ = − 046	95% CI [− 570, 499]	*p* = 862
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition versus color naming (condition 3 minus condition 1, scaled scores)	10 4 (2 2)	17	ρ = − 584	95% CI [− 904, − 028]	*p* = 014

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Table 3 (*continued*)

Citation and year	Language test(s) were administered in	Self-report test	*M* (*SD*)	Neuropsychological test	*M* (*SD*)	*N* in correlation	Raw correlation	95% CI	*p* value
		BRIEF-A/shift subscale[b]		D-KEFS/Inhibition/shifting versus color naming (condition 4 minus condition 1, scaled scores)	10 8 (2 0)	17	ρ = − 498	95% CI [− 849, 049]	*p* = 042
Meltzer et al (2017)[a]		BRIEF-A/shift subscale	9 6 (2 3)	Stroop/Golden Interference Score[b,c]	38 5 (8 1)	46	*r* = 262	95% CI [− 073, 538]	*p* = 079
		BRIEF-A/shift subscale		Stroop/Color–Word Trial	25 3 (11 1)	47	*r* = 234	95% CI [− 023, 503]	*p* = 113
		BRIEF-A/shift subscale		TMT/Part B time	161 0 (62 9)	41	*r* = 033	95% CI [− 303, 333]	*p* = 837
		BRIEF-A/shift subscale		TMT/Part B-A time[b,c]	100 8 (50 2)	41	*r* = − 094	95% CI [− 386, 226]	*p* = 559
		BRIEF-A/shift subscale		TMT/Part B/A time[b,c]	2 8 (0 8)	41	*r* = − 385	95% CI [− 577, − 142]	*p* = 013
		BRIEF-A/shift subscale		COWAT/total number correct	25 2 (10 5)	51	ρ = − 260	95% CI [− 473, 005]	*p* = 065
Okuda et al (2017)[a]	CFS, Stroop, WCST, TMT; Japanese versions								
Preintervention									
		CFS/total score	33 6 (11 3)	WCST/% Perseverative errors	1 4 (2 2)	5	*r* = 683	95% CI [612, 1 00]	*p* = 204
		CFS/total score		Stroop/Interference Score/ trial 3 − trial 1	4 9 (7 2)	5	*r* = 985	95% CI [889, 1 00]	*p* = 002
		CFS/total score		TMT/Part B time[b]	62 9 (13 1)	5	*r* = 662	95% CI [−1 00, 1 00]	*p* = 224
		CFS/total score		TMT/Part B-A time[c]	25 1 (7 2)	5	*r* = 367	95% CI [−1 00, 1 00]	*p* = 543
		CFS/total score		TMT/Part B/A time[b,c]	1 7 (0 2)	5	*r* = − 027	95% CI [−1 00, 1 00]	*p* = 966
		CFS/total score		The Brixton Spatial Anticipation Test/Total errors	15 0 (14 6)	5	*r* = 450	95% CI [−1 00, 1 00]	*p* = 447
Postintervention									
		CFS/total score	41 0 (7 4)	Stroop/Interference Score/ trial 3 − trial 1	4 2 (3 9)	5	*r* = 799	95% CI [−1 00, 1 00]	*p* = 105
		CFS/total score		TMT/Part B time[b]	43 3 (7 5)	5	*r* = 658	95% CI [− 296, 1 00]	*p* = 228
		CFS/total score		TMT/Part B-A time[c]	20 3 (6 8)	5	*r* = 919	95% CI [482, 1 00]	*p* = 027
		CFS/total score		TMT/Part B/A time[b,c]	1 9 (0 4)	5	*r* = 885	95% CI [− 553, 1 00]	*p* = 046
		CFS/total score		The Brixton Spatial Anticipation Test/Total errors	6 4 (3 9)	5	*r* = 701	95% CI [408, 1 00]	*p* = 187
Follow-up									
		CFS/total score	36 2 (11 9)	WCST/% Perseverative errors	0 4 (0 5)	5	*r* = 138	95% CI [−1 00, 1 00]	*p* = 825
		CFS/total score		Stroop/Interference Score/ trial 3 − trial 1	4 5 (3 7)	5	*r* = 872	95% CI [− 506, 1 00]	*p* = 054
		CFS/total score		TMT/Part B time[b]	42 6 (9 0)	5	ρ = 671	95% CI [− 304, 1 00]	*p* = 215
		CFS/total score		TMT/Part B-A time[c]	24 7 (17 5)	5	ρ = 894	95% CI [725, 1 00]	*p* = 041
		CFS/total score		TMT/Part B/A time[b,c]	5 6 (8 4)	5	ρ = 894	95% CI [725, 1 00]	*p* = 041
		CFS/total score		The Brixton Spatial Anticipation Test/Total errors	8 4 (5 7)	5	*r* = 578	95% CI [− 019, 1 00]	*p* = 308

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Table 3 (*continued*)

Citation and year	Language test(s) were administered in	Self-report test	*M* (*SD*)	Neuropsychological test	*M* (*SD*)	*N* in correlation	Raw correlation	95% CI	*p* value
Whiting et al (2015)[a]		AAQ-ABI	25 6 (7 8)	TMT/Part B time (*z*-score)	−2 5 (3 7)	50	ρ = 262	95% CI [− 043, 520]	*p* = 066
		AAQ-ABI		TMT/Part B-A time (*z*-score)[b,c]	−1 4 (2 5)	50	ρ = 219	95% CI [− 108, 524]	*p* = 127
		AAQ-II	18 7 (9 9)	TMT/Part B time (*z*-score)		50	ρ = − 103	95% CI [− 409, 205]	*p* = 478
		AAQ-II		TMT/Part B-A time (*z*-score)[b,c]		50	ρ = − 103	95% CI [− 457, 203]	*p* = 476

Note *M* = mean; *N* = sample size; CI = confidence interval; *p* = *p* value; BRIEF-A = Behavior Rating Inventory of Executive Function—Adult; TMT = Trail Making Test; *r* = Pearson's correlation coefficient; WCST = Wisconsin Card Sorting Test; ρ = Spearman's correlation coefficient; D-KEFS = Delis–Kaplan Executive Function System; CFI = Cognitive Flexibility Inventory; COWAT = Controlled Oral Word Association Test; TAU = treatment as usual; CFS = Cognitive Flexibility Scale; CR = cognitive remediation; AAQ-ABI = Acceptance and Action Questionnaire—Acquired Brain Injury; AAQ-II = Acceptance and Action Questionnaire-II

[a] Reported data do not match the data reported in the original article in order to fulfill the eligibility criteria [b] Additional outcomes that were obtained from the authors (i e , not reported in the original citation) [c] Direction of test scores (i e , higher/lower scores indicate better/worse performance) was unable to be confirmed

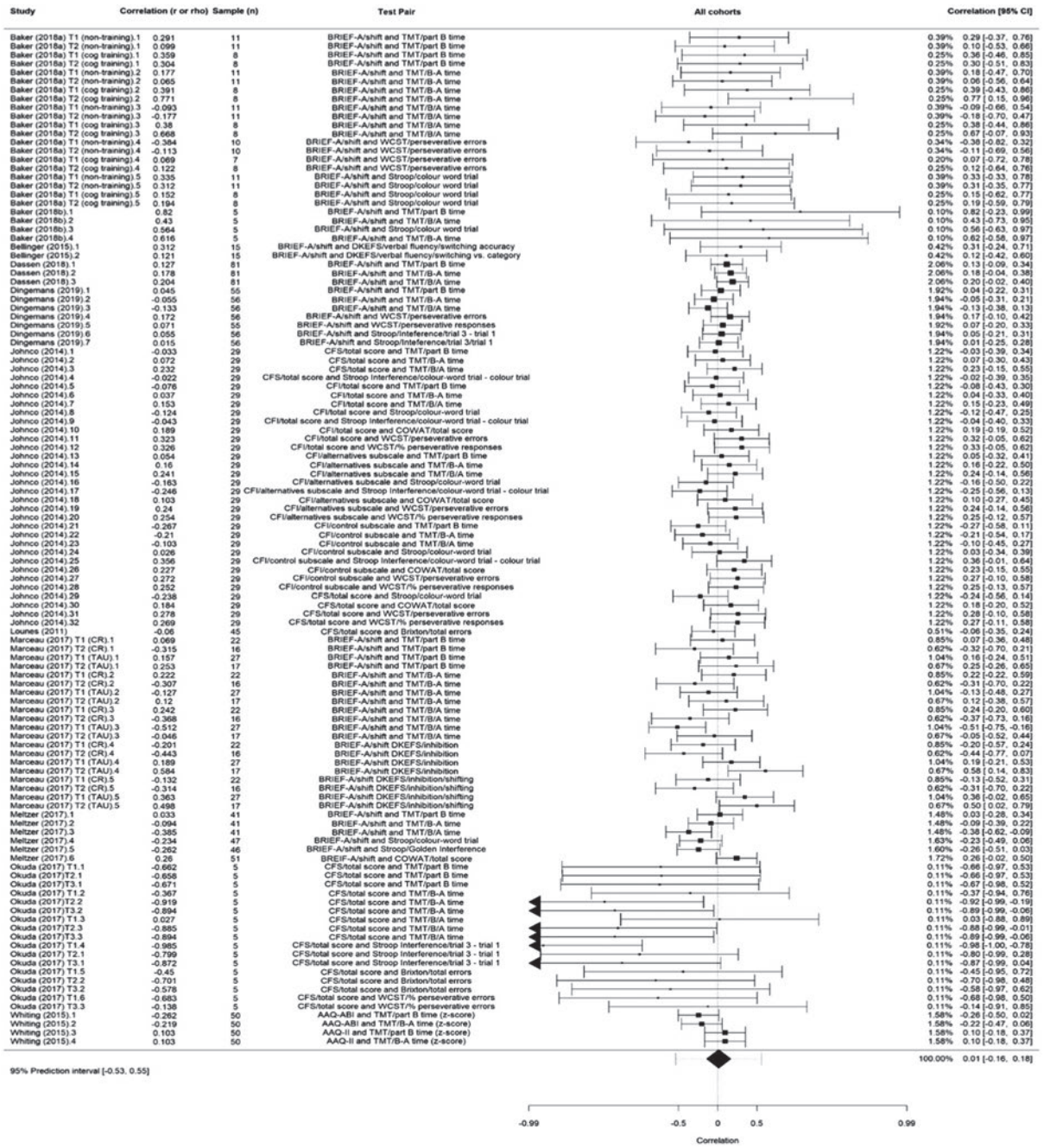
(which we anticipated would influence self report and neuropsychological tests equally) were not excluded. In both Baker et al. (2018a) and Marceau et al. (2017), we noticed between group variability for particular test pairs at baseline, for people with chronic back pain and substance use disorder, respectively. That is, for Baker et al. (2018a), the active control group (nontraining) had a negative correlation, while the intervention group (cognitive training) had a positive correlation for the BRIEF A shift subscale and the WCST (perseverative errors), and the BRIEF A shift subscale and the TMT B/A. Marceau et al. (2017) displayed the same between group variability for the BRIEF A shift subscale and the TMT B A, and the BRIEF A shift subscale and the TMT B/A. That is, the control group (treatment as usual) had a negative correlation for the two aforementioned test pairs at baseline, while the intervention group (cognitive remediation) had a positive correlation at baseline. For the BRIEF A shift subscale and the D KEFS (inhibition minus color naming score), and the BRIEF A shift and D KEFS (inhibition/shifting minus color naming score), Marceau et al. (2017) displayed the reversed pattern such that the intervention group had a negative correlation and the control group had a positive correlation. To summarize, even at baseline, there was variability between intervention and control groups as well as in the size and direction of correlations between self report and neuropsychological tests. More importantly, the direction varied depending on which test pairs were considered, even though all the tests were intended to detect the same construct (i.e., cognitive flexibility). The most parsimonious interpretation of these variations is that performance on these tests is vulnerable to extraneous factors or these tests are simply unreliable. Although neuropsychological tests are generally considered to be more robust assessments of "cognitive flexibility" than self reported approaches and thus, are accepted as the "gold standard" (Johnco et al., 2014a; Lounes et al., 2011), reliability data are limited for either type of test. Even when previous literature has found favorable reliability coefficients for the neuropsychological tests such as the WCST (Loeber et al., 2009; Miranda et al., 2020), comparisons across studies are confounded by differences in scoring procedures as well as differences in test modality (e.g., computerized vs. manual administration). Further, neuropsychological tests, particularly the WCST, are largely susceptible to practice effects upon reassessment (Basso et al., 1999, 2001), which potentially may have contributed to the observed changes, but are not likely to fully explain them.

Risk of Bias

An important consideration is the methodological quality of the included studies. All 11 included studies were assessed as having a moderate to high level of bias and therefore, considered to be of poor quality. Methodological and reporting issues were common potential sources of bias in the constituent papers. It is noteworthy that none of the included studies preregistered a protocol, despite the fundamental role they play in promoting transparent reporting, reproducibility, and replicability within psychological sciences (Klein et al., 2018) and other relevant fields (Lee et al., 2018). Additionally, none of the studies were adequately powered for correlational analyses, and only two studies (Dingemans et al., 2019; Okuda et al., 2017) reported a priori power calculations; however, Okuda et al. (2017) did not reach their required sample and for Dingemans et al. (2019), their minimum effect of interest was

Figure 4

Multilevel Random Effects Meta Analysis: Correlations of All Test Pairs Accounting for Dependency of Effect Sizes



Note. AAQ ABI Acceptance and Action Questionnaire Acquired Brain Injury; AAQ II Acceptance and Action Questionnaire II; BRIEF A shift Behavior Rating Inventory of Executive Function Adult (shift subscale); Brixton The Brixton Spatial Anticipation Test; CFI Cognitive Flexibility Inventory; CFS Cognitive Flexibility Scale; CI confidence interval; COWAT Controlled Oral Word Association Test; CR cognitive remediation; D KEFS Delis Kaplan Executive Function System; *n* sample size; *r* Pearson's correlation coefficient; ρ Spearman's correlation coefficient; TAU treatment as usual; T1 time point 1; T2 time point 2; TMT Trail Making Test; WCST Wisconsin Card Sorting Test.

unclear, increasing the risk that statistically significant results observed in some of our included studies may potentially be false positive findings. The limited power of constituent studies to detect relationships may contribute to the resilience of the field's understanding that a relationship exists. That is, perhaps keen observers of the literature who have not detected a relationship, may conclude that: "a lack of evidence for a relationship is not evidence against a relationship." Our review clearly counters that

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

position because we did not find a lack of evidence; we found no relationship in either our multilevel, nor individual random effects meta analyses. We acknowledge that our largest individual random effects meta analyses involved a total of 205 participants, which leaves open the possibility that very small relationships between particular pairs of tests, may still exist. A larger number of included studies would of course increase our power, but the precision estimates were not excessive, suggesting that if a relationship does indeed exist, it is likely to be too small to offer any practical utility. Whether or not our meta analyses lacked statistical power remains uncertain we did not perform a priori power calculations for our individual random effects meta analysis, and methods for power calculations for multilevel random effects meta analyses are debated and statistical software is currently unavailable (Griffin, 2021).

Strengths and Limitations

This review has particular strengths and limitations. We prospectively registered the current review on PROSPERO, registration number: CRD42019131567; available from (https://www.crd.york.ac.uk/prospero/display record.php?ID=CRD42019131567), and OSF (https://osf.io/mxjfu/). All deviations from our original protocol were recorded and clearly noted. We followed the PRISMA guidelines (Moher et al., 2009) and at least two independent reviewers were involved in all stages of the review process: article screening, data extraction, risk of bias assessments, and data analysis. No limits were placed on the search strategy. That is, we did not limit the search by language, country of origin, date of publication, or publication status, making our search strategy comprehensive. However, we chose to confine our review to studies with participants over 18 years of age because adolescence is a period of vulnerability and rapid neurocognitive and physical change, often associated with the development of psychopathologies and during which the brain circuitry thought to be important for flexibility is still undergoing development (Uddin, 2021). Our approach to contacting authors was also thorough, such that we contacted authors via email, phone, and ResearchGate (ResearchGate GmbH, Berlin, Germany), enabling us to obtain raw or aggregate data from a total of 11 authors.

The large heterogeneity of test pairs and clinical conditions presented in this review meant that it was not possible to pool studies that involved participants with the same clinical diagnosis. With the exception of back pain (Baker et al., 2018a, 2018b), there were no two studies that had sufficient data for the same clinical group. We include data from a range of clinical diagnoses and there does not seem to be diagnosis specific results. As such, generalization beyond these cohorts should consider that on the one hand, the broad picture of the data suggests that it is unlikely that results from these two assessment approaches would relate in any particular cohort, but on the other hand, that the available data are not exhaustive. Further to this, Baker et al. (2018a), Okuda et al. (2017), and Marceau et al. (2017) each had multiple observations for a particular test pair, but they were obtained from the same individuals so performing a meta analysis would have been methodologically flawed. In some instances, we were unable to determine the direction of the test scores (i.e., whether higher/lower scores indicate better/worse performance). In these cases, we assumed the direction of the test scores that would be predicted (i.e., positive or negative correlation), an approach that would increase the risk of a false positive finding, which was clearly not an issue here. In October 2020, we made a necessary deviation from our protocol when we reran the database searches, but not the gray literature searches. We took this approach because we found that, with the gray literature, determining study eligibility invariably required finding and contacting authors, which was a very laborious and time consuming step that we considered to offer little chance of affecting our results. We therefore decided that for our updated search, additional studies would only be included if the eligibility could be determined without contacting authors. We recognize that it is possible that we missed papers because of this strategy, but we contend that it is unlikely that we missed papers that would overturn our main findings.

We included studies irrespective of their primary aim or research question and did not formally assess publication bias. Although we cannot exclude such a bias, Lange (2020) has previously argued that including primary studies that address different research questions to that of a meta analysis may be less susceptible to publication bias because its publication status is not dependent on aspects related to the research question (e.g., statistical significance and originality). Additionally, that we did not identify a relationship between self report and neuropsychological tests of cognitive flexibility (a result that contradicts the false positive findings observed with the "file drawer problem" whereby null findings are less likely to be published), makes it unlikely that publication bias explains our results. We included studies that (a) used a test that was already used as an assessment of cognitive flexibility (determined a priori) or (b) the authors stated a priori that the test was being used as an assessment of cognitive flexibility. This put the onus on the authors of primary studies to be satisfied that the tests were valid assessments of cognitive flexibility, which may have led to the inclusion of tests that are not valid, and the exclusion of tests that are. We did not independently evaluate bespoke tests and instead relied on authors and their peer review to judge the validity of the tests that were implemented. Examining whether such tests were consistent with the authors operationalized definition of what constitutes a test of cognitive flexibility may have added value; however, we saw no difference in results between more widely recognizable and commonly used tests and less common or bespoke assessments.

Across the 11 included articles, 12 different tests were used as assessments of cognitive flexibility. Moreover, individual studies used different versions of a single neuropsychological task (e.g., paper and pencil vs. computerized versions of the TMT) and across studies, there was a lack of consistency regarding which outcome variables were used as indices of cognitive flexibility (particularly for the TMT). In fact, two large systematic reviews investigating set shifting difficulties in people with chronic pain (Berryman et al., 2014; Bunk et al., 2019) reported similar findings in that a variety of different tasks and outcome variables were used in their retrieved studies. The large variability in outcome measures poses significant challenges when attempting to compare findings across studies (Lang et al., 2014) and as such, previous studies have been unable to perform meta analyses as a result of this variability (Miles et al., 2020). To address and overcome this problem, we requested various outcome measures of tests from authors (e.g., TMT part B and TMT B/A scores). Therefore, meta analyses were only possible when at least two authors (a) used the same test pair (b) provided

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

additional outcome variables to allow for a specific test pair or (c) provided the raw data to allow for specific variables to be calculated.

A key issue identified in this review was the inconsistency in how the outcome variables of the TMT are reported. Part B of the TMT is considered to be an assessment of set shifting ability, while part A assesses visual search and psychomotor speed capacity (Vall & Wade, 2015). This review found that 45% of studies did not control for baseline visual search and psychomotor speed (TMT part A), which can be calculated by subtracting the time taken to complete part A (in seconds) from the time taken to complete part B. Removing the influence of visual search and psychomotor speed is supposed to provide an index of cognitive flexibility. In fact, some of the included studies reported only part B, while others reported part A and part B independently. However, relying on part B as an outcome of cognitive flexibility is problematic because performance will also depend on visual search and psychomotor speed. Equally problematic is the fact that the reliability of difference scores such as that of the TMT B A are considerably low, making such scores inadequate for clinical interpretations (Kopp, 2011). Although the TMT B/A score has also been validated as an indicator of cognitive flexibility, with scores greater than 3 reflecting impaired performance (Arbuthnott & Frank, 2000), that validation study was undertaken in a group of university undergraduates, and thus whether the results hold up in clinical samples remains unknown.

Recommendations for Research and Clinical Practice

This review raises important theoretical and practical implications for our field. Perhaps most importantly, self report and neuropsychological tests of "cognitive flexibility" do not appear to assess the same underlying latent construct, which requires the field to rethink what in fact they do assess. Our findings challenge these traditional notions and thus, the results prompt reassessment of prior research in our field. It seems imperative that our field also reconsider the terminology used to describe these tests, to avoid using them interchangeably, and move away from language that conflates their descriptions and toward more precise terms, for example, "processing efficiency tests" (i.e., relying on the algorithmic mind). It may be important to also acknowledge that self report tests depend on aptitude to self reflect (i.e., relying on the reflective mind) in our discussion of self report tools. These suggestions have already been made for performance based and self rating tests of so called "executive function" more broadly (Stanovich, 2009; Toplak et al., 2013).

An avenue for future research is to establish the psychometric properties (e.g., construct, convergent and discriminant validity, and internal reliability) of outcome variables on various self report and neuropsychological tests across clinically, culturally, and age diverse groups. Although previous studies have found moderate to nonexistent correlations between accepted outcomes of "cognitive flexibility" on the WCST and TMT (Chaytor et al., 2006; Herbrich et al., 2019; Kortte et al., 2002; O'Donnell et al., 1994; Van Autreve et al., 2013), the studies vary in the outcome variables, versions, and modality of the tests they use. Such issues limit comparisons across studies, make it difficult to draw robust conclusions about the validity of the test outcomes, and also make it challenging to clarify what construct these neuropsychological tests are ultimately assessing. Further, it is important to note that a portion of these studies report variables that may not be an appropriate index of "cognitive flexibility" (i.e., TMT part B; Vall & Wade, 2015 or TMT B A; Kopp, 2011). It may be prudent to start by validating the most widely acknowledged outcomes of "cognitive flexibility" (i.e., perseverative errors/responses of the WCST; Miles et al., 2021) and aim to confirm the findings through adequately powered replication studies. As stated by the Open Science Collaboration, "direct replication provides the opportunity to assess and improve reproducibility" (Open Science Collaboration, 2015, p. 2). Hence, one possible solution is for researchers to take part in crowd sourced replication studies (Nosek et al., 2012), which can only lead to improved scientific understanding of the psychometric properties and core constructs that underpin such assessments. Alternatively, we can aim to use improved versions of tests that more specifically target "cognitive flexibility." For example, previous versions of the TMT such as the original version (Reitan, 1958), the D KEFS version (Delis et al., 2001), or the Comprehensive TMT (Reynolds, 2002) do not produce adequate reliability coefficients for outcomes of cognitive flexibility, however, the Brunswick version of the TMT (Kopp et al., 2008) appears to be a more reliable and valid assessment of the construct because of its increased trial lengths and attempts to better separate core components of the task: Visual search, visuomotor and cognitive flexibility.

Similar problems exist for self report tests of "cognitive flexibility." The original CFI (Dennis & Vander Wal, 2010) has been translated into various languages such as Mandarin (Wang et al., 2016), Turkish (Sapmaz & Doğan, 2013), Russian (Kurginyan & Osavolyuk, 2018), Italian (Portoghese et al., 2020), and Persian (Shareh et al., 2014), while the original CFS (Martin & Rubin, 1995) has been translated into Japanese (Oshiro et al., 2016) and Turkish (Çelikkaleli, 2014), with both tests demonstrating favorable psychometric properties including, but not limited to, internal consistency (Kurginyan & Osavolyuk, 2018; Maltby et al., 2004; Martin & Anderson, 1998; Oshiro et al., 2016; Portoghese et al., 2020; Shareh et al., 2014), test retest reliability (Dennis & Vander Wal, 2010; Kurginyan & Osavolyuk, 2018; Martin & Rubin, 1995; Oshiro et al., 2016; Shareh et al., 2014), and convergent construct validity (Dennis & Vander Wal, 2010; Johnco et al., 2014a; Portoghese et al., 2020). It is noteworthy that all of the above mentioned studies with the exception of Johnco et al. (2014a) used student samples, limiting the generalisability of the findings. Further, the exact psychometric properties investigated are not always consistent across studies, which pose significant challenges when comparing findings. Another concern is that some studies have been unable to replicate the original factor structure of the CFI (Kurginyan & Osavolyuk, 2018; Portoghese et al., 2020; Shareh et al., 2014). Scores on self report tests of "cognitive flexibility" have been negatively associated with depressive and anxiety symptomology (Dennis & Vander Wal, 2010; Johnco et al., 2014a; Lounes et al., 2011; Wang et al., 2019; Yu et al., 2020) and perfectionism (Çutuk, 2020; Hayatbini et al., 2021), which may be potential confounding factors that could explain the inconsistencies reported in previous findings. Therefore, future research using self report and neuropsychological tests of "cognitive flexibility" should aim to control for such confounding variables.

Our findings emphasize that both researchers and clinicians should also be cautious about implying that a specific neuropsychological test assesses "cognitive flexibility." Even when two tests share good convergent validity, it does not necessarily mean that

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

both tests are targeting "cognitive flexibility" per se. Part of the reason for this may be because of the aforementioned task impurity problem (Buchsbaum et al., 2005; Miyake et al., 2000; Wildes et al., 2014). Therefore, future research should acknowledge that the exact mechanisms responsible for neuropsychological task performance, and the processes that underpin the apparent inflexibility of behaviors observed in those with clinical conditions, remain to be elucidated.

Future research would benefit from registration of protocols prior to commencing data collection an established practice for clinical trials and recommended in psychological (Klein et al., 2018; Nosek et al., 2012, 2018) and related (Lee et al., 2018) sciences clear declarations of funding, and potential or actual conflicts of interest. A priori sample size calculations and declaration of effect size of interest, using reliable tests or verifying the reliability of new tests, counterbalancing the order of assessments in cases of multiple assessments, and avoiding overcapitalizing on chance by limiting the number of outcomes and analyses, will all improve the confidence we can have in experimental results. Methodological improvements specific to "cognitive flexibility" research include being aware of the shortcomings of various outcomes of the TMT and using more reliable and valid versions where possible, reporting the language, version, and modality of tests used, and using the most targeted outcome from tests with many outcomes, for example, the WCST.

Constraints on Generality

The included studies largely consisted of a WEIRD sample (Western, Educated, Industrialized, Rich, and Democratic; Henrich et al., 2010), and thus the extent to which the results are generalizable to populations with different socioeconomical, political, and educational backgrounds is unknown. Further, we believe our findings will be reproducible with the clinical cohorts presented in this review, however, it is noteworthy that we cannot generalize our results to all clinical cohorts. As cognitive flexibility follows a unique trajectory in early development, replication studies or systematic inquiries asking the same research question as ours are required to elucidate the extent of relationship between self report and neuropsychological tests of cognitive flexibility in children and adolescents. We have no reason to suspect that the results depend on other characteristics of the participants, materials, procedures, or contextual factors.

Conclusion

We aimed to determine the relationship between self report and neuropsychological tests of cognitive flexibility in clinical cohorts. Our findings contrast the notion that the two approaches assess the same construct they clearly do not. Our findings should compel both researchers and clinicians to avoid using self report and neuropsychological tests of "cognitive flexibility" interchangeably as evaluative tools in clinical populations, or inferring or predicting performance on one approach on the basis of performance on the other. Our review identified substantial gaps in the knowledge base with respect to methodological rigor, reporting, and interpretation.

References

Abbate Daga, G., Buzzichelli, S., Marzola, E., Amianto, F., & Fassino, S. (2014). Clinical investigation of set shifting subtypes in anorexia nervosa. *Psychiatry Research*, *219*(3), 592 597. https://doi.org/10.1016/j.psychres.2014.06.024

Albein Urios, N., Youssef, G. J., Kirkovski, M., & Enticott, P. G. (2018). Autism spectrum traits linked with reduced performance on self report behavioural measures of cognitive flexibility. *Journal of Autism and Developmental Disorders*, *48*(7), 2506 2515. https://doi.org/10.1007/s10803 018 3503 3

Alfimova, M. V., Golimbet, V., Lezheiko, T., & Korovaitseva, G. (2017). A potential role of the 5 HTTLPR polymorphism in self reported executive functioning. *The Spanish Journal of Psychology*, *20*(e 13), Article E13. https://doi.org/10.1017/sjp.2017.6

Arbuthnott, K., & Frank, J. (2000). Trail making test, part B as a measure of executive control: Validation using a set switching paradigm. *Journal of Clinical and Experimental Neuropsychology*, *22*(4), 518 528. https://doi.org/10.1076/1380 3395(200008)22:4;1 0;FT518

Arlt, J., Yiu, A., Eneva, K., Taylor Dryman, M., Heimberg, R. G., & Chen, E. Y. (2016). Contributions of cognitive inflexibility to eating disorder and social anxiety symptoms. *Eating Behaviors*, *21*, 30 32. https://doi.org/10.1016/j.eatbeh.2015.12.008

Ashendorf, L., & McCaffrey, R. J. (2008). Exploring age related decline on the Wisconsin card sorting test. *The Clinical Neuropsychologist*, *22*(2), 262 272. https://doi.org/10.1080/13854040701218436

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three level meta analytic models in R: A step by step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154 174. https://doi.org/10.20982/tqmp.12.3.p154

Aytug, Z. G., Rua, T., Brazeal, D. V., Almaraz, J. A., & González, C. B. (2018). A socio cultural approach to multicultural experience: Why interactions matter for creative thinking but exposures don't. *International Journal of Intercultural Relations*, *64*, 29 42. https://doi.org/10.1016/j.ijintrel.2018.03.004

Baker, K. S., Georgiou Karistianis, N., Lampit, A., Valenzuela, M., Gibson, S. J., & Giummarra, M. J. (2018a). Computerised training improves cognitive performance in chronic pain: A participant blinded randomised active controlled trial with remote supervision. *Pain*, *159*(4), 644 655. https://doi.org/10.1097/j.pain.0000000000001150

Baker, K. S., Gibson, S. J., Georgiou Karistianis, N., & Giummarra, M. J. (2018b). Relationship between self reported cognitive difficulties, objective neuropsychological test performance and psychological distress in chronic pain. *European Journal of Pain (London, England)*, *22*(3), 601 613. https://doi.org/10.1002/ejp.1151

Ballantyne, E. C. (2013). *The neuropsychology of mindfulness: Attentional aspects* (Publication No. 3578540) [Doctoral dissertation]. Adler School of Professional Psychology. ProQuest Dissertations and Theses Global.

Baniqued, P. L., Kranz, M. B., Voss, M. W., Lee, H., Cosman, J. D., Severson, J., & Kramer, A. F. (2014). Cognitive training with casual video games: Points to consider. *Frontiers in Psychology*, *4*, Article 1010. https://doi.org/10.3389/fpsyg.2013.01010

Bannon, E. E. (2017). *The efficacy of acceptance based behavior therapy versus cognitive therapy for test anxiety and working memory performance* (Publication No. 10754246) [Doctoral dissertation]. Bowling Green State University. ProQuest Dissertations and Theses Global.

Banz, B. C. (2014). *An evaluation of executive functions, cognitive control and a neurocognitive profile of college binge* (Publication No. 3635581) [Doctoral dissertation]. Colorado State University. ProQuest Dissertations and Theses Global.

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *13*(3), 283 292. https://doi.org/10.1076/clin.13.3.283.1743

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Basso, M. R., Lowery, N., Ghormley, C., & Bornstein, R. A. (2001). Practice effects on the wisconsin card sorting test 64 card version across 12 months. *The Clinical Neuropsychologist*, *15*(4), 471 478. https://doi.org/10.1076/clin.15.4.471.1883

Bell, T., Trost, Z., Buelow, M. T., Clay, O., Younger, J., Moore, D., & Crowe, M. (2018). Meta analysis of cognitive performance in fibromyal gia. *Journal of Clinical and Experimental Neuropsychology*, *40*(7), 698 714. https://doi.org/10.1080/13803395.2017.1422699

Bellinger, D. C., Watson, C. G., Rivkin, M. J., Robertson, R. L., Roberts, A. E., Stopp, C., Dunbar Masterson, C., Bernson, D., DeMaso, D. R., Wypij, D., & Newburger, J. W. (2015). Neuropsychological status and structural brain imaging in adolescents with single ventricle who under went the fontan procedure. *Journal of the American Heart Association*, *4*(12), Article e002302. https://doi.org/10.1161/JAHA.115.002302

Benton, L. A., Hamsher, K., & Sivan, A. B. (1994). Controlled oral word association test. *Multilingual aphasia examination* (3rd ed.). AJA Associates.

Berryman, C., Stanton, T. R., Bowering, K. J., Tabor, A., McFarlane, A., & Moseley, G. L. (2014). Do people with chronic pain have impaired executive function? A meta analytical review. *Clinical Psychology Review*, *34*(7), 563 579. https://doi.org/10.1016/j.cpr.2014.08.003

Bettis, A. H., Coiro, M. J., England, J., Murphy, L. K., Zelkowitz, R. L., Dejardins, L., Eskridge, R., Adery, L. H., Yarboi, J., Pardo, D., & Compas, B. E. (2017). Comparison of two approaches to prevention of mental health problems in college students: Enhancing coping and executive function skills. *Journal of American College Health*, *65*(5), 313 322. https://doi.org/10.1080/07448481.2017.1312411

Biederman, J., Fried, R., Tarko, L., Surman, C., Spencer, T., Pope, A., Grossman, R., McDermott, K., Woodworth, K. Y., & Faraone, S. V. (2017). Memantine in the treatment of executive function deficits in adults with ADHD: A pilot randomized double blind controlled clinical trial. *Journal of Attention Disorders*, *21*(4), 343 352. https://doi.org/10.1177/1087054714538656

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire II: A revised measure of psychological inflexibility and experiential avoidance. *Behav ior Therapy*, *42*(4), 676 688. https://doi.org/10.1016/j.beth.2011.03.007

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta analysis*. Wiley. https://doi.org/10.1002/9780470743386

Boschloo, A., Krabbendam, L., Aben, A., de Groot, R., & Jolles, J. (2014). Sorting test, tower test, and BRIEF SR do not predict school performance of healthy adolescents in preuniversity education. *Frontiers in Psychol ogy*, *5*(287), Article 287. https://doi.org/10.3389/fpsyg.2014.00287

Brenner, L. (2017). *The relationship between measures of cognitive flexibil ity and intimate partner violence perpetration in a court mandated domestic violence treatment program* (Publication No. 10264484) [Doc toral dissertation]. Rosalind Franklin University of Medicine and Science. ProQuest Dissertations and Theses Global.

Brockmeyer, T., Ingenerf, K., Walther, S., Wild, B., Hartmann, M., Herzog, W., Bents, H., & Friederich, H. C. (2014). Training cognitive flexibility in patients with anorexia nervosa: A pilot randomized controlled trial of cognitive remediation therapy. *International Journal of Eating Disorders*, *47*(1), 24 31. https://doi.org/10.1002/eat.22206

Buchsbaum, B. R., Greer, S., Chang, W. L., & Berman, K. F. (2005). Meta analysis of neuroimaging studies of the Wisconsin card sorting task and component processes. *Human Brain Mapping*, *25*(1), 35 45. https://doi.org/10.1002/hbm.20128

Bunk, S., Preis, L., Zuidema, S., Lautenbacher, S., & Kunz, M. (2019). Executive functions and pain. *Journal of Neuropsychology*, *30*(3), 169 196. https://doi.org/10.1024/1016 264X/a000264

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton tests*. Thames Valley Test Company.

Burmester, B., Leathem, J., & Merrick, P. (2016). Subjective cognitive complaints and objective cognitive function in aging: A systematic review and meta analysis of recent cross sectional findings. *Neuropsychology Review*, *26*(4), 376 393. https://doi.org/10.1007/s11065 016 9332 2

Cañas, J., Quesada, J. F., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem solving tasks. *Ergonomics*, *46*(5), 482 501. https://doi.org/10.1080/0014013031000061640

Çelikkaleli, Ö. (2014). The validity and reliability of the cognitive flexibility scale. *Education in Science*, *39*(176), 339 346. https://doi.org/10.15390/EB.2014.3466

Chamberlain, S. R., Fineberg, N. A., Menzies, L. A., Blackwell, A. D., Bullmore, E. T., Robbins, T. W., & Sahakian, B. J. (2007). Impaired cognitive flexibility and motor inhibition in unaffected first degree rela tives of patients with obsessive compulsive disorder. *The American Journal of Psychiatry*, *164*(2), 335 338. https://doi.org/10.1176/ajp.2007.164.2.335

Chaytor, N., Schmitter Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*(3), 217 227. https://doi.org/10.1016/j.acn.2005.12.002

Clayson, K. A. (2018). *A multimethod examination of the relevance of executive control to disgust and mental contamination among female assault survivors* (Publication No. 106173522018) [Doctoral dissertation] Baylor University. ProQuest Dissertations and Theses Global.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155 159. https://doi.org/10.1037/0033 2909.112.1.155

Cohen, J. N., Jensen, D., Stange, J. P., Neuburger, M., & Heimberg, R. G. (2017). The immediate and long term effects of an intensive meditation retreat. *Mindfulness*, *8*(4), 1064 1077. https://doi.org/10.1007/s12671 017 0682 5

Coleman, S. L. (2010). *The relationship between cognitive inhibition and intrusive thoughts* (Publication No. 3423272) [Doctoral dissertation] State University of New York at Binghamtom. ProQuest Dissertations and Theses Global.

Çutuk, Z. A. (2020). Relationship between multidimensional perfectionism and cognitive flexibility in university students. *Journal of Human Sciences*, *17*(4), 1263 1274. https://doi.org/10.14687/jhs.v17i4.6090

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571 578. https://doi.org/10.1016/j.tins.2015.07.003

Daly, B. P., Hildenbrand, A. K., Turner, E., Berkowitz, S., & Tarazi, R. A. (2017). Executive functioning among college students with and without history of childhood maltreatment. *Journal of Aggression, Maltreatment and Trauma*, *26*(7), 717 735. https://doi.org/10.1080/10926771.2017.1317685

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self report and behavioural measures weakly correlated? *Trends in Cognitive Sciences*, *24*(4), 267 269. https://doi.org/10.1016/j.tics.2020.01.007

Dassen, F. C. M., Houben, K., Allom, V., & Jansen, A. (2018). Self regulation and obesity: The role of executive function and delay discount ing in the prediction of weight loss. *Journal of Behavioral Medicine*, *41*(6), 806 818. https://doi.org/10.1007/s10865 018 9940 9

Dawson, E. L., Shear, P. K., & Strakowski, S. M. (2012). Behavior regulation and mood predict social functioning among healthy young adults. *Journal of Clinical and Experimental Neuropsychology*, *34*(3), 297 305. https://doi.org/10.1080/13803395.2011.639297

Deák, G. O. (2004). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior*, *31*, 271 327. https://doi.org/10.1016/S0065 2407(03)31007 9

Del Re, A. C. (2015). A practical tutorial on conducting meta analysis in R. *The Quantitative Methods for Psychology*, *11*(1), 37 50. https://doi.org/10.20982/tqmp.11.1.p037

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis Kaplan executive functioning system*.

Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis Kaplan executive function system: An update. *Journal of the International Neuropsychological Society*, *10*(2), 301 303. https://doi.org/10.1017/S1355617704102191

Demetriou, E. A., Song, C. Y., Park, S. H., Pepper, K. L., Naismith, S. L., Hermens, D. F., Hickie, I. B., Thomas, E. E., Norton, A., White, D., & Guastella, A. J. (2018). Autism, early psychosis, and social anxiety disorder: A transdiagnostic examination of executive function cognitive circuitry and contribution to disability. *Translational Psychiatry*, *8*(1), Article 200. https://doi.org/10.1038/s41398 018 0193 8

Dennis, J. P., & Vander Wal, J. S. (2010). The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. *Cogni tive Therapy and Research*, *34*(3), 241 253. https://doi.org/10.1007/s10608 009 9276 4

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135 168. https://doi.org/10.1146/annurev psych 113011 143750

Dickson, K. S., Ciesla, J. A., & Zelic, K. (2017). The role of executive functioning in adolescent rumination and depression. *Cognitive Therapy and Research*, *41*, 62 72. https://doi.org/10.1007/s10608 016 9802 0

Dingemans, A. E., Vanhaelen, C. B., Aardoom, J. J., & van Furth, E. F. (2019). The influence of depressive symptoms on executive functioning in binge eating disorder: A comparison of patients and non obese healthy controls. *Psychiatry Research*, *274*, 138 145. https://doi.org/10.1016/j.psychres.2019.02.033

Donders, J., Oh, Y. I., & Gable, J. (2015). Self and informant ratings of executive functioning after mild traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, *30*(6), E30 E39. https://doi.org/10.1097/HTR.0000000000000120

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross sectional studies (AXIS). *BMJ Open*, *6*(12), Article e011458. https://doi.org/10.1136/bmjopen 2016 011458

Drake, C. E. (2008). *Evaluating a behavioral measure of psychological flexibility* (Publication No. 3329528) [Doctoral dissertation]. The Univer sity of Mississippi. ProQuest Dissertations and Theses Global.

Dumbrava, I. L. (2017). An empirical approach to wisdom processes. *Romanian Journal of Applied Psychology*, *19*(2), 50 58. https://doi.org/10.24913/rjap.19.2.04

Eliava, N. (1964). *A problem of set in cognitive psychology*. Academy of Science Press.

Follmer, D. J., & Stefanou, C. R. (2014). Examining the correspondence between a direct and an indirect measure of executive functions: Implica tions for school based assessment. *The School Psychologist*, *68*(3), 12 18.

Fujino, J., Tei, S., Itahashi, T., Aoki, Y., Ohta, H., Kubota, M., Isobe, M., Hashimoto, R. I., Nakamura, M., Kato, N., & Takahashi, H. (2019). Need for closure and cognitive flexibility in individuals with autism spectrum disorder: A preliminary study. *Psychiatry Research*, *271*, 247 252. https://doi.org/10.1016/j.psychres.2018.11.057

Fujino, J., Tei, S., Jankowski, K. F., Kawada, R., Murai, T., & Takahashi, H. (2017). Role of spontaneous brain activity in explicit and implicit aspects of cognitive flexibility under socially conflicting situations: A resting state fMRI study using fractional amplitude of low frequency fluctuations. *Neuroscience*, *367*, 60 71. https://doi.org/10.1016/j.neuroscience.2017.10.025

Gajewski, P. D., Ferdinand, N. K., Kray, J., & Falkenstein, M. (2018). Understanding sources of adult age differences in task switching: Evidence from behavioral and ERP studies. *Neuroscience and Biobehav ioral Reviews*, *92*, 255 275. https://doi.org/10.1016/j.neubiorev.2018.05.029

García, A. V., Torrecillas, F. L., de Arcos, F. A., & García, M. P. (2005). Effects of executive impairments on maladaptive explanatory styles in substance abusers: Clinical implications. *Archives of Clinical Neuropsy chology*, *20*(1), 67 80. https://doi.org/10.1016/j.acn.2004.03.002

García Molina, A., Tormos, J. M., Bernabeu, M., Junqué, C., & Roig Rovira, T. (2012). Do traditional executive measures tell us anything about daily life functioning after traumatic brain injury in Spanish speaking individuals? *Brain Injury*, *26*(6), 864 874. https://doi.org/10.3109/02699052.2012.655362

Garcia Willingham, N. E., Roach, A. R., Kasarskis, E. J., & Segerstrom, S. C. (2018). Self regulation and executive functioning as related to survival in motor neuron disease: Preliminary findings. *Psychosomatic Medicine*, *80*(7), 665 672. https://doi.org/10.1097/PSY.0000000000000602

Gelonch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and objective measures. *Comprehensive Psychiatry*, *66*, 113 122. https://doi.org/10.1016/j.comppsych.2016.01.002

Genders, R., & Tchanturia, K. (2010). Cognitive remediation therapy (CRT) for anorexia in group format: A pilot study. *Eating and Weight Disorders*, *15*(4), e234 e239. https://doi.org/10.1007/BF03325304

Geurts, H. M., Corbett, B., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in Cognitive Sciences*, *13*(2), 74 82. https://doi.org/10.1016/j.tics.2008.11.006

Golden, C. J. (1975). A group version of the stroop color and word test. *Journal of Personality Assessment*, *39*(4), 386 388. https://doi.org/10.1207/s15327752jpa3904 10

Gonzalez, C. A., Figueroa, I. J., Bellows, B. G., Rhodes, D., & Youmans, R. J. (2013). A new behavioral measure of cognitive flexibility. In D. Harris (Ed.), *International conference on engineering psychology and cognitive ergonomics* (pp. 297 306). Springer. https://doi.org/10.1007/978 3 642 39360 0

Good, D. (2014). Predicting real time adaptive performance in a dynamic decision making context. *Journal of Management and Organization*, *20*(6), 715 732. https://doi.org/10.1017/jmo.2014.54

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl type card sorting problem. *Journal of Experimental Psychology*, *38*(4), 404 411. https://doi.org/10.1037/h0059831

Griffin, J. W. (2021). Calculating statistical power for meta analysis using metapower. *The Quantitative Methods for Psychology*, *17*(1), 24 39. https://doi.org/10.20982/tqmp.17.1.p024

Guy, S. C., Isquith, P. K., & Gioia, G. A. (2004). *BRIEF SR: Behavior rating inventory of executive function self report version: Professional manual*. Psychological Assessment Resources.

Hagen, E., Erga, A. H., Hagen, K. P., Nesvåg, S. M., McKay, J. R., Lundervold, A. J., & Walderhaug, E. (2016). Assessment of executive function in patients with substance use disorder: A comparison of inven tory and performance based assessment. *Journal of Substance Abuse Treatment*, *66*, 1 8. https://doi.org/10.1016/j.jsat.2016.02.010

Hanssen, K. T., Beiske, A. G., Landrø, N. I., & Hessen, E. (2014). Predictors of executive complaints and executive deficits in multiple sclerosis. *Acta Neurologica Scandinavica*, *129*(4), 234 242. https://doi.org/10.1111/ane.12177

Hayatbini, N., Knauft, K., & Kalia, V. (2021). Cognitive reappraisal moderates the relationship between perfectionism and cognitive flexibil ity. *Journal of Clinical Psychology*, *77*, 1685 1699. https://doi.org/10.1002/jclp.23124

Healy, B., Treadwell, A., & Reagan, M. (2011). Measures of RSA suppres sion, attentional control, and negative affect predict self ratings of execu tive functions. *Journal of Psychophysiology*, *25*(4), 164 173. https://doi.org/10.1027/0269 8803/a000053

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test manual: Revised and expanded*. Psychological Assessment Resources.

Heinonen, K., Pesonen, A. K., Lahti, J., Pyhälä, R., Strang Karlsson, S., Hovi, P., Järvenpää, A. L., Eriksson, J. G., Andersson, S., Kajantie, E., &

Raikkonen, K. (2013). Self and parent rated executive functioning in young adults with very low birth weight. *Pediatrics*, *131*(1), e243 e250. https://doi.org/10.1542/peds.2012 0839

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), Article 29. https://doi.org/10.1038/466029a

Herbrich, L. R., Kappel, V., Winter, S. M., & van Noort, B. M. (2019). Executive functioning in adolescent anorexia nervosa: Neuropsychology versus self and parental report. *Child Neuropsychology*, *25*(6), 816 835. https://doi.org/10.1080/09297049.2018.1536200

Hertzberg, L. C. (2017). *The differential roles of objective neuropsychologi cal testing and self report measures in assessing neurocognitive impairment in relation to addiction severity among alcohol abusing individuals* (Publication No. 10287393) [Doctoral dissertation]. Baylor University. ProQuest Dissertations and Theses Global.

Higgins, J. P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J. , & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). Wiley. https://doi.org/10.1002/ 9781119536604

Hocking, M. C., Hobbie, W. L., Deatrick, J. A., Lucas, M. S., Szabo, M. M., Volpe, E. M., & Barakat, L. P. (2011). Neurocognitive and family functioning and quality of life among young adult survivors of childhood brain tumors. *The Clinical Neuropsychologist*, *25*(6), 942 962. https:// doi.org/10.1080/13854046.2011.580284

Holliday, J., Tchanturia, K., Landau, S., Collier, D., & Treasure, J. (2005). Is impaired set shifting an endophenotype of anorexia nervosa? *The Ameri can Journal of Psychiatry*, *162*(12), 2269 2275. https://doi.org/10.1176/ appi.ajp.162.12.2269

Howlett, C. A., Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., Karran, E. L., Szeto, K., Pronk, L., Miles, S., & Moseley, G. L. (2021). Same room different windows? A systematic review and meta analysis of the relationship between self report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, *88*, Article 102061. https://doi.org/10.1016/j.cpr.2021.102061

IBM Corp. (2011). *IBM SPSS statistics for windows* (Version 20.0.) [Com puter software]. https://www.ibm.com/au en/analytics/spss statistics software

Inozemtseva, O., Pérez Solís, L., Matute, E., & Juárez, J. (2016). Differ ential improvement of executive functions during abstinence in cocaine dependent patients: A longitudinal study. *Substance Use and Misuse*, *51*(11), 1428 1440. https://doi.org/10.1080/10826084 .2016.1178293

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, *30*(2), 190 200. https://doi.org/10.1016/j.newideapsych .2011.11.001

Ionescu, T. (2017). The variability stability flexibility pattern: A possible key to understadning the flexibility of the human mind. *Review of General Psychology*, *21*(2), 123 131. https://doi.org/10.1037/gpr0000110

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014a). Reliability and validity of two self report measures of cognitive flexibility. *Psychological Assessment*, *26*(4), 1381 1387. https://doi.org/10.1037/a0038009

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014b). The influence of cognitive flexibility on treatment outcome and cognitive restructuring skill acquisition during cognitive behavioural treatment for anxiety and depres sion in older adults: Results of a pilot study. *Behaviour Research and Therapy*, *57*, 55 64. https://doi.org/10.1016/j.brat.2014.04.005

Kalkut, E. (2010). *Neuropsychological assessment of executive functioning and its association with depressive symptomology* (Publication No. 3419840) [Doctoral dissertation]. Loyola University Chicago. ProQuest Dissertations and Theses Global.

Kesler, S., Hadi Hosseini, S. M., Heckler, C., Janelsins, M., Palesh, O., Mustian, K., & Morrow, G. (2013). Cognitive training for improving executive function in chemotherapy treated breast cancer survivors. *Clin ical Breast Cancer*, *13*(4), 299 306. https://doi.org/10.1016/j.clbc.2013 .02.004

Kesler, S. R., Kent, J. S., & O'Hara, R. (2011). Prefrontal cortex and executive function impairments in primary breast cancer. *Archives of Neurology*, *68*(11), 1447 1453. https://doi.org/10.1001/archneurol.2011.245

Kim, H. Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry and Endodontics*, *38*(1), 52 54. https://doi.org/10.5395/rde.2013.38.1.52

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsonne, G., Vanpaemel, W., Frank, M. C., Nuijten, M., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra. Psychology*, *4*(1), 2 15. https://doi.org/10.1525/colla bra.158

Kopp, B. (2011). Neuropsychologists must keep their eyes on the reliability of difference measures. *Journal of the International Neuropsychological Society*, *17*(3), 562 563. https://doi.org/10.1017/S1355617711000361

Kopp, B., Rösser, N., & Wessel, K. (2008). Psychometric characteristics and practice effects of the Brunswick trail making test. *Perceptual and Motor Skills*, *107*(3), 707 733. https://doi.org/10.2466/pms.107.3 .707 733

Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The trail making test, part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, *9*(2), 106 109. https://doi.org/10.1207/S15324826A N0902 5

Kurginyan, S. S., & Osavolyuk, E. Y. (2018). Psychometric properties of a Russian version of the Cognitive Flexibility Inventory (CFI R). *Frontiers in Psychology*, *9*, Article 845. https://doi.org/10.3389/fpsyg.2018.00845

Kuula, L., Pesonen, A. K., Heinonen, K., Kajantie, E., Eriksson, J. G., Andersson, S., Lano, A., Lahti, J., Wolke, D., & Räikkönen, K. (2018). Naturally occurring circadian rhythm and sleep duration are related to executive functions in early adulthood. *Journal of Sleep Research*, *27*(1), 113 119. https://doi.org/10.1111/jsr.12581

Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Controlled Clinical Trials*, *2*(2), 93 113. https:// doi.org/10.1016/0197 2456(81)90001 5

Lang, K., Lloyd, S., Khondoker, M., Simic, M., Treasure, J., & Tchanturia, K. (2015). Do children and adolescents with anorexia nervosa display an inefficient cognitive processing style? *PLOS ONE*, *10*(7), Article e0131724. https://doi.org/10.1371/journal.pone.0131724

Lang, K., Stahl, D., Espie, J., Treasure, J., & Tchanturia, K. (2014). Set shifting in children and adolescents with anorexia nervosa: An exploratory systematic review and meta analysis. *International Journal of Eating Disorders*, *47*(4), 394 399. https://doi.org/10.1002/eat.22235

Lange, F. (2020). Are difficult to study populations too difficult to study in a reliable way: Lessons learned from meta analyses in clinical neuropsy chology. *European Psychologist*, *25*(1), 41 50. https://doi.org/10.1027/ 1016 9040/a000384

Lao Kaim, N. P., Fonville, L., Giampietro, V. P., Williams, S. C. R., Simmons, A., & Tchanturia, K. (2015). Aberrant function of learning and cognitive control networks underlie inefficient cognitive flexibility in anorexia nervosa: A cross sectional fMRI study. *PLOS ONE*, *10*(5), Article e0124027. https://doi.org/10.1371/journal.pone.0124027

Lau, K. (2012). *The role of cognitive flexibility in acculturation and bicultural stress among Chinese American adults* (Publication No. 3517364) [Doctoral dissertation]. Alliant International University. Pro Quest Dissertations and Theses Global.

Lee, H., Lamb, S. E., Bagg, M. K., Toomey, E., Cashin, A. G., & Moseley, G. L. (2018). Reproducible and replicable pain research: A critical review. *Pain*, *159*(9), 1683 1689. https://doi.org/10.1097/j.pain.0000000000001254

Lee, J. (1999). *Social intelligence: Flexibility and complexity are distinct from creativity* (Publication No. 9924645) [Doctoral dissertation]. Uni versity of Notre Dame. ProQuest Dissertations and Theses Global.

Lee, R. S., Hermens, D. F., Porter, M. A., & Redoblado Hodge, M. A. (2012). A meta analysis of cognitive deficits in first episode major depressive disorder. *Journal of Affective Disorders*, *140*(2), 113 124. https://doi.org/10.1016/j.jad.2011.10.023

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Leppanen, J., Adamson, J., & Tchanturia, K. (2018). Impact of cognitive remediation therapy on neurocognitive processing in anorexia nervosa. *Frontiers in Psychiatry*, *9*, Article 13. https://doi.org/10.3389/fpsyt.2018.00096

Li, Y., Dong, F., Cao, F., Cui, N., Li, J., & Long, Z. (2013). Poly victimization and executive functions in junior college students. *Scandinavian Journal of Psychology*, *54*(6), 485 492. https://doi.org/10.1111/sjop.12083

Lieberman, L., Gorka, S. M., Sarapas, C., & Shankman, S. A. (2016). Cognitive flexibility mediates the relation between intolerance of uncertainty and safety signal responding in those with panic disorder. *Cognition and Emotion*, *30*(8), 1495 1503. https://doi.org/10.1080/02699931.2015.1067189

Loeber, S., Duka, T., Welzel, H., Nakovics, H., Heinz, A., Flor, H., & Mann, K. (2009). Impairment of cognitive abilities and decision making after chronic use of alcohol: The impact of multiple detoxifications. *Alcohol and Alcoholism*, *44*(4), 372 381. https://doi.org/10.1093/alcalc/agp030

Lounes, N., Khan, G., & Tchanturia, K. (2011). Assessment of cognitive flexibility in anorexia nervosa Self report or experimental measure? A brief report. *Journal of the International Neuropsychological Society*, *17*(5), 925 928. https://doi.org/10.1017/S1355617711000671

Løvstad, M., Funderud, I., Endestad, T., Due Tønnessen, P., Meling, T. R., Lindgren, M., Knight, R. T., & Solbakk, A. K. (2012a). Executive functions after orbital or lateral prefrontal lesions: Neuropsychological profiles and self reported executive functions in everyday living. *Brain Injury*, *26*(13 14), 1586 1598. https://doi.org/10.3109/02699052.2012.698787

Løvstad, M., Funderud, I., Meling, T., Krämer, U. M., Voytek, B., Due Tønnessen, P., Endestad, T., Lindgren, M., Knight, R. T., & Solbakk, A. K. (2012b). Anterior cingulate cortex and cognitive control: Neuropsychological and electrophysiological findings in two patients with lesions to dorsomedial prefrontal cortex. *Brain and Cognition*, *80*(2), 237 249. https://doi.org/10.1016/j.bandc.2012.07.008

Lüdecke, D. (2019). *esc: Effect size computation for meta analysis (version 0.5.1)*. https://CRAN.R project.org/package esc

Maltby, J., Day, L., McCutcheon, L. E., Martin, M. M., & Cayanus, J. L. (2004). Celebrity worship, cognitive flexibility, and social complexity. *Personality and Individual Differences*, *37*(7), 1475 1482. https://doi.org/10.1016/j.paid.2004.02.004

Mangum, R. W. (2018). *Self understanding of executive functions in individuals with agenesis of the corpus callosum* (Publication No. 10792935) [Doctoral dissertation]. Fuller Theological Seminary, School of Psychology. ProQuest Dissertations and Theses Global.

Marceau, E. M., Berry, J., Lunn, J., Kelly, P. J., & Solowij, N. (2017). Cognitive remediation improves executive functions, self regulation and quality of life in residents of a substance use disorder therapeutic community. *Drug and Alcohol Dependence*, *178*, 150 158. https://doi.org/10.1016/j.drugalcdep.2017.04.023

Marceau, E. M., Kelly, P. J., & Solowij, N. (2018). The relationship between executive functions and emotion regulation in females attending therapeutic community treatment for substance use disorder. *Drug and Alcohol Dependence*, *182*, 58 66. https://doi.org/10.1016/j.drugalcdep.2017.10.008

Martin, M. M., & Anderson, C. M. (1998). The cognitive flexibility scale: Three validity studies. *International Journal of Phytoremediation*, *11*(1), 1 9. https://doi.org/10.1080/08934219809367680

Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports*, *76*(2), 623 626. https://doi.org/10.2466/pr0.1995.76.2.623

McAnarney, E. R., Zarcone, J., Singh, P., Michels, J., Welsh, S., Litteer, T., Wang, H., & Klein, J. D. (2011). Restrictive anorexia nervosa and set shifting in adolescents: A biobehavioral interface. *The Journal of Adolescent Health*, *49*(1), 99 101. https://doi.org/10.1016/j.jadohealth.2010.11.259

McCloskey, L. (2010). Response frequency and Exner's active passive ratio as Rorschach indicators of flexibility. *Perceptual and Motor Skills*, *111*(2), 559 575. https://doi.org/10.2466/02.03.04.09.PMS.111.5.559 575

McCurdy, M. D., Turner, E. M., Barakat, L. P., Hobbie, W. L., Deatrick, J. A., Paltin, I., Fisher, M. J., & Hocking, M. C. (2016). Discrepancies among measures of executive functioning in a subsample of young adult survivors of childhood brain tumor: Associations with treatment intensity. *Journal of the International Neuropsychological Society*, *22*(9), 900 910. https://doi.org/10.1017/S1355617716000771

Mckowen, J., Carrellas, N., Zulauf, C., Ward, E. N., Fried, R., & Wilens, T. (2017). Factors associated with attrition in substance using patients enrolled in an intensive outpatient program. *American Journal on Addictions*, *26*(8), 780 787. https://doi.org/10.1111/ajad.12619

McKowen, J. W., Isenberg, B. M., Carrellas, N. W., Zulauf, C. A., Ward, N. E., Fried, R. S., & Wilens, T. E. (2018). Neuropsychological changes in patients with substance use disorder after completion of a one month intensive outpatient treatment program. *American Journal on Addictions*, *27*(8), 632 638. https://doi.org/10.1111/ajad.12824

Meltzer, E. P., Kapoor, A., Fogel, J., Elbulok Charcape, M. M., Roth, R. M., Katz, M. J., Lipton, R. B., & Rabin, L. A. (2017). Association of psychological, cognitive, and functional variables with self reported executive functioning in a sample of nondemented community dwelling older adults. *Applied Neuropsychology. Adult*, *24*(4), 364 375. https://doi.org/10.1080/23279095.2016.1185428

Miles, S., Gnatt, I., Phillipou, A., & Nedeljkovic, M. (2020). Cognitive flexibility in acute anorexia nervosa and after recovery: A systematic review. *Clinical Psychology Review*, *81*, Article 101905. https://doi.org/10.1016/j.cpr.2020.101905

Miles, S., Howlett, C. A., Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the Wisconsin card sorting test to assess cognitive flexibility. *Behavior Research Methods*, *53*, 2083 2091. https://doi.org/10.3758/s13428 021 01551 3

Miranda, A. R., Franchetto Sierra, J., Martínez Roulet, A., Rivadero, L., Serra, S. V., & Soria, E. A. (2020). Age, education and gender effects on Wisconsin card sorting test: Standardization, reliability and validity in healthy Argentinian adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *27*(6), 807 825. https://doi.org/10.1080/13825585.2019.1693491

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49 100. https://doi.org/10.1006/cogp.1999.0734

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097. https://doi.org/10.1371/journal.pmed.1000097

Müller, U., & Kerns, K. (2015). The development of executive function. In L. S. Liben, U. Müller, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (pp. 571 623). Wiley. https://doi.org/10.1002/9781118963418.childpsy214

National Heart, Lung, and Blood Insitute. (2014). *Quality assessment tool for observational cohort and cross sectional studies*. https://www.nhlbi.nih.gov/health topics/study quality assessment tools

Nguyen, L., Murphy, K., & Andrews, G. (2019). Cognitive and neural plasticity in old age: A systematic review of evidence from executive functions cognitive training. *Ageing Research Reviews*, *53*, Article 100912. https://doi.org/10.1016/j.arr.2019.100912

Nordvall, O., Jonsson, B., & Neely, A. S. (2017). Self reported and performance based measures of executive functions in interned youth. *Psychology, Crime and Law*, *23*(3), 240 253. https://doi.org/10.1080/1068316X.2016.1239725

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600 2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615 631. https://doi.org/10.1177/1745691612459058

O'Donnell, J. P., Macgregor, L. A., Dabrowski, J. J., Oestreicher, J. M., & Romero, J. J. (1994). Construct validity of neuropsychological tests of conceptual and attentional abilities. *Journal of Clinical Psychology*, *50*(4), 596 600. https://doi.org/10.1002/1097 4679(199407)50:4<596::AID JCLP2270500416>3.0.CO;2 S

Oguz, G., Celikbas, Z., Batmaz, S., Cagli, S., & Sungur, M. Z. (2019). Comparison between obsessive compulsive disorder and panic disorder on metacognitive beliefs, emotional schemas, and cognitive flexibility. *International Journal of Cognitive Therapy*, *12*(3), 157 178. https://doi.org/10.1007/s41811 019 00047 5

Okuda, T., Asano, K., Numata, N., Hirano, Y., Yamamoto, T., Tanaka, M., Matsuzawa, D., Shimizu, E., Iyo, M., & Nakazato, M. (2017). Feasibility of cognitive remediation therapy for adults with autism spectrum disorders: A single group pilot study. *Neuropsychiatric Disease and Treatment*, *13*, 2185 2191. https://doi.org/10.2147/NDT.S141555

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), Article 30. https://doi.org/10.1126/science.aac4716

Oshiro, K., Nagaoka, S., & Shimizu, E. (2016). Development and validation of the Japanese version of cognitive flexibility scale. *BMC Research Notes*, *9*(1), Article 275. https://doi.org/10.1186/s13104 016 2070 y

Oswald, A. (2015). *Effects of repeated exposure to stressors on cognitive flexibility as a function of self type and self complexity: A multi session experimental test* [Unpublished Doctoral dissertation]. University of Newcastle.

Palm, K. M. (2004). *An examination of mindfulness: Assessment and relationship to PTSD* (Publication No. 3161689) [Doctoral dissertation]. University of Nevada, Reno. ProQuest Dissertations and Theses Global.

Pang, E. W., Dunkley, B. T., Doesburg, S. M., da Costa, L., & Taylor, M. J. (2016). Reduced brain connectivity and mental flexibility in mild traumatic brain injury. *Annals of Clinical and Translational Neurology*, *3*(2), 124 131. https://doi.org/10.1002/acn3.280

Perkins, S. C., Cortina, K. S., Smith Darden, J. P., & Graham Bermann, S. A. (2012). The mediating role of self regulation between intrafamilial violence and mental health adjustment in incarcerated male adolescents. *Journal of Interpersonal Violence*, *27*(7), 1199 1224. https://doi.org/10.1177/0886260511425240

Perkins, S. C., Smith Darden, J., Ametrano, R. M., & Graham Bermann, S. (2014). Typologies of violence exposure and cognitive processing in incarcerated male adolescents. *Journal of Family Violence*, *29*(4), 439 451. https://doi.org/10.1007/s10896 014 9600 y

Pope, C. N., Ross, L. A., & Stavrinos, D. (2016). Association between executive function and problematic adolescent driving. *Journal of Developmental and Behavioral Pediatrics*, *37*(9), 702 711. https://doi.org/10.1097/DBP.0000000000000353

Portoghese, I., Lasio, M., Conti, R., Mascia, M. L., Hitchcott, P., Agus, M., Gemignani, A., & Penna, M. P. (2020). Cognitive Flexibility Inventory: Factor structure, invariance, reliability, convergent, and discriminant validity among Italian university students. *PsyCh Journal*, *9*(6), 934 941. https://doi.org/10.1002/pchj.401

Rabin, L. A., Roth, R. M., Isquith, P. K., Wishart, H. A., Nutter Upham, K. E., Pare, N., Flashman, L., & Saykin, A. J. (2006). Self and informant reports of executive function on the BRIEF A in MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology*, *21*(7), 721 732. https://doi.org/10.1016/j.acn.2006.08.004

R Core Team. (2019). *R: A language and environment for statistical computing (Version 3.6.2)* [Computer software]. R Foundation for Statistical Computing. https://www.r project.org/

Reed, R. G., Combs, H. L., & Segerstrom, S. C. (2020). The structure of self regulation and its psychological and physical health correlates in older adults. *Collabra. Psychology*, *6*(1), Article 23. https://doi.org/10.1525/collabra.297

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271 276. https://doi.org/10.2466/pms.1958.8.3.271

Renwick, B., Musiat, P., Lose, A., DeJong, H., Broadbent, H., Kenyon, M., Loomes, R., Watson, C., Ghelani, S., Serpell, L., Richards, L., Johnson Sabine, E., Boughton, N., Treasure, J., & Schmidt, U. (2015). Neuro and social cognitive clustering highlights distinct profiles in adults with anorexia nervosa. *International Journal of Eating Disorders*, *48*(1), 26 34. https://doi.org/10.1002/eat.22366

Resnik, D. B. (2007). Conflicts of interest in scientific research related to regulation or litigation. *The Journal of Philosophy, Science & Law*, *7*(1), 1 12. https://doi.org/10.5840/jpsl2007722

Reynolds, C. R. (2002). *Comprehensive trail making test (CTMT): Examiner's manual*. PRO ED.

Roberts, M. E., Tchanturia, K., Stahl, D., Southgate, L., & Treasure, J. (2007). A systematic review and meta analysis of set shifting ability in eating disorders. *Psychological Medicine*, *37*(8), 1075 1084. https://doi.org/10.1017/S0033291707009877

Roberts, M. E., Tchanturia, K., & Treasure, J. L. (2010). Exploring the neurocognitive signature of poor set shifting in anorexia and bulimia nervosa. *Journal of Psychiatric Research*, *44*(14), 964 970. https://doi.org/10.1016/j.jpsychires.2010.03.001

Rouel, M., Raman, J., Hay, P., & Smith, E. (2016). Validation of the behaviour rating inventory of executive function Adult version (BRIEF A) in the obese with and without binge eating disorder. *Eating Behaviors*, *23*, 58 65. https://doi.org/10.1016/j.eatbeh.2016.07.010

Sakamoto, M., Woods, S. P., Kolessar, M., Kriz, D., Anderson, J. R., Olavarria, H., Sasaki, A. W., Chang, M., Flora, K. D., Loftis, J. M., & Huckans, M. (2013). Protective effects of higher cognitive reserve for neuropsychological and daily functioning among individuals infected with hepatitis C. *Journal of Neurovirology*, *19*(5), 442 451. https://doi.org/10.1007/s13365 013 0196 4

Salmon, D. (1999). *Attentional control and the deconstruction of the schemas underlying the perception of pain* (Publication No. 9934176) [Doctoral dissertation]. Long Island University. ProQuest Dissertations and Theses Global.

Sapmaz, F., & Doğan, T. (2013). Assessment of cognitive flexibility: Reliability and validity studies of Turkish version of the Cognitive Flexibility Inventory. *Journal of Faculty of Educational Sciences*, *46*(1), 143 161. https://doi.org/10.1501/egifak 0000001278

Saylik, R., & Szameitat, A. J. (2018). The association between negative attributional style and working memory performance. *The Open Psychology Journal*, *11*, 131 141. https://doi.org/10.2174/1874350101811010131

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, *7*, Article 16. https://doi.org/10.1186/1472 6947 7 16

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & the PRISMA P Group. (2015). Preferred reporting items for systematic review and meta analysis protocols (PRISMA P) 2015: Elaboration and explanation. *BMJ (Clinical Research Ed.)*, *350*, Article g7647. https://doi.org/10.1136/bmj.g7647

Shareh, H., Farmani, A., & Soltani, E. (2014). Investigating the reliability and validity of the Cognitive Flexibility Inventory (CFI I) among Iranian university students. *Practice in Clinical Psychology*, *2*(1), 43 50.

Shuster, J., & Toplak, M. E. (2009). Executive and motivational inhibition: Associations with self report measures related to inhibition.

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

Consciousness and Cognition, *18*(2), 471 480. https://doi.org/10.1016/j.concog.2009.01.004

Sølsnes, A. E., Skranes, J., Brubakk, A. M., & Løhaugen, G. C. C. (2014). Executive functions in very low birth weight young adults: A comparison between self report and neuropsychological test results. *Journal of the International Neuropsychological Society*, *20*(5), 506 515. https://doi.org/10.1017/S1355617714000332

Spencer, R. J., Drag, L. L., Walker, S. J., & Bieliauskas, L. A. (2010). Self reported cognitive symptoms following mild traumatic brain injury are poorly associated with neuropsychological performance in OIF/OEF veterans. *Journal of Rehabilitation Research and Development*, *47*(6), 521 530. https://doi.org/10.1682/JRRD.2009.11.0181

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri process theory? In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55 88). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230167.001.0001

Sternheim, L., van der Burgh, M., Berkhout, L. J., Dekker, M. R., & Ruiter, C. (2014). Poor cognitive flexibility, and the experience thereof, in a subclinical sample of female students with obsessive compulsive symptoms. *Scandinavian Journal of Psychology*, *55*(6), 573 577. https://doi.org/10.1111/sjop.12163

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643 662. https://doi.org/10.1037/h0054651

Suchy, Y., Turner, S. L., Queen, T. L., Durracio, K., Wiebe, D. J., Butner, J., Franchow, E. I., White, P. C., Murray, M. A., Swinyard, M., & Berg, C. A. (2016). The relation of questionnaire and performance based measures of executive functioning with type 1 diabetes outcomes among late adolescents. *Health Psychology*, *35*(7), 661 669. https://doi.org/10.1037/hea0000326

Svendsen, T. S., Erga, A. H., Hagen, E., McKay, J. R., Njå, A. L. M., Årstad, J., & Nesvåg, S. (2017). How to maintain high retention rates in long term research on addiction: A case report. *Journal of Social Work Practice in the Addictions*, *17*(4), 374 387. https://doi.org/10.1080/1533256X.2017.1361831

Sylvester, M. (2011). *Acceptance and Commitment Therapy for improving adaptive functioning in persons with a history of pediatric acquired brian injury* (Publication No. 3472789) [Doctoral dissertation]. University of Nedada. ProQuest Dissertations and Theses Global.

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: A first report. *Journal of the American Academy of Audiology*, *24*(7), 616 634. https://doi.org/10.3766/jaaa.24.7.10

Tchanturia, K., Anderluh, M. B., Morris, R. G., Rabe Hesketh, S., Collier, D. A., Sanchez, P., & Treasure, J. L. (2004). Cognitive flexibility in anorexia nervosa and bulimia nervosa. *Journal of the International Neuropsychological Society*, *10*(4), 513 520. https://doi.org/10.1017/S1355617704104086

Tchanturia, K., Davies, H., & Campbell, I. C. (2007). Cognitive remediation therapy for patients with anorexia nervosa: Preliminary findings. *Annals of General Psychiatry*, *6*(1), Article 199. https://doi.org/10.1186/1744859X 6 14

Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., Treasure, J., Morris, R., & Morris, R. (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PLOS ONE*, *7*(1), Article e28331. https://doi.org/10.1371/journal.pone.0028331

Tchanturia, K., Larsson, E., & Brown, A. (2016). Benefits of group cognitive remediation therapy in anorexia nervosa: Case series. *Neuropsychiatrie (Deisenhofen)*, *30*(1), 42 49. https://doi.org/10.1007/s40211 016 0177 y

Tchanturia, K., Morris, R. G., Surguladze, S., & Treasure, J. (2002). An examination of perceptual and cognitive set shifting tasks in acute anorexia nervosa and following recovery. *Eating and Weight Disorders*, *7*(4), 312 315. https://doi.org/10.1007/BF03324978

Tchanturia, K., Serpell, L., Troop, N., & Treasure, J. (2001). Perceptual illusions in eating disorders: Rigid and fluctuating styles. *Journal of Behavior Therapy and Experimental Psychiatry*, *32*(3), 107 115. https://doi.org/10.1016/S0005 7916(01)00025 8

Tenconi, E., Santonastaso, P., Degortes, D., Bosello, R., Titton, F., Mapelli, D., & Favaro, A. (2010). Set shifting abilities, central coherence, and handedness in anorexia nervosa patients, their unaffected siblings and healthy controls: Exploring putative endophenotypes. *World Journal of Biological Psychiatry*, *11*(6), 813 823. https://doi.org/10.3109/15622975.2010.483250

The Cochrane Collaboration. (2014). *Review manager (RevMan) (Version 5.3)* [Computer software]. The Nordic Cochrane Centre. https://training.cochrane.org/

Thoma, P., Wiebel, B., & Daum, I. (2007). Response inhibition and cognitive flexibility in schizophrenia with and without comorbid substance use disorder. *Schizophrenia Research*, *92*(1 3), 168 180. https://doi.org/10.1016/j.schres.2007.02.004

Thompson, R. (2010). *Understanding bulimia nervosa from a neuropsychological perspective: Impulsivity and binge purge behavior in adolescent and young adult women* (Publication No. 3372002) [Doctoral dissertation]. Temple University. ProQuest Dissertations and Theses Global.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(2), 131 143. https://doi.org/10.1111/jcpp.12001

Tornås, S., Løvstad, M., Solbakk, A. K., Evans, J., Endestad, T., Hol, P. K., Schanke, A. K., & Stubberud, J. (2016). Rehabilitation of executive functions in patients with chronic acquired brain injury with goal management training, external cuing, and emotional regulation: A randomized controlled trial. *Journal of the International Neuropsychological Society*, *22*(4), 436 452. https://doi.org/10.1017/S1355617715001344

Uddin, L. Q. (2021). Cognitive and behavioural flexibility: Neural mechanisms and clinical considerations. *Nature Reviews Neuroscience*, *22*(3), 167 179. https://doi.org/10.1038/s41583 021 00428 w

Uznadze, D. (1966). *The psychology of set*. Consultants Bureau.

Vall, E., & Wade, T. D. (2015). Trail making task performance in inpatients with anorexia nervosa and bulimia nervosa. *European Eating Disorders Review*, *23*(4), 304 311. https://doi.org/10.1002/erv.2364

Van Autreve, S., De Baene, W., Baeken, C., van Heeringen, C., & Vervaet, M. (2013). Do restrictive and bingeing/purging subtypes of anorexia nervosa differ on central coherence and set shifting? *European Eating Disorders Review*, *21*(4), 308 314. https://doi.org/10.1002/erv.2233

van Noort, B. M., Kraus, M. K. A., Pfeiffer, E., Lehmkuhl, U., & Kappel, V. (2016). Neuropsychological and behavioural short term effects of cognitive remediation therapy in adolescent anorexia nervosa: A pilot study. *European Eating Disorders Review*, *24*(1), 69 74. https://doi.org/10.1002/erv.2383

van Zellem, L., Buysse, C., Madderom, M., Legerstee, J. S., Aarsen, F., Tibboel, D., & Utens, E. M. (2015). Long term neuropsychological outcomes in children and adolescents after cardiac arrest. *Intensive Care Medicine*, *41*(6), 1057 1066. https://doi.org/10.1007/s00134 015 3789 y

Viechtbauer, W. (2010). Conducting meta analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1 48. https://doi.org/10.18637/jss.v036.i03

Wang, T., Li, M., Xu, S., Liu, B., Wu, T., Lu, F., Xie, J., Peng, L., & Wang, J. (2019). Relations between trait anxiety and depression: A mediated moderation model. *Journal of Affective Disorders*, *244*, 217 222. https://doi.org/10.1016/j.jad.2018.09.074

Wang, Y., Yang, Y., Xiao, W. T., & Su, Q. (2016). Validity and reliability of the Chinese version of the Cognitive Flexibility Inventory in college students. *Chinese Mental Health Journal*, *30*(1), 58 62.

Warren, S. L., Crocker, L. D., Spielberg, J. M., Engels, A. S., Banich, M. T., Sutton, B. P., Miller, G. A., & Heller, W. (2013). Cortical organization of inhibition related functions and modulation by psychopathology. *Frontiers in Human Neuroscience*, *7*, Article 271. https://doi.org/10.3389/fnhum.2013.00271

Wenner, C. J., Bianchi, J., Figueredo, A. J., Rushton, J. P., & Jacobs, W. J. (2013). Life history theory and social deviance: The mediating role of executive function. *Intelligence*, *41*(2), 102 113. https://doi.org/10.1016/j.intell.2012.11.004

Westwood, H., Mandy, W., & Tchanturia, K. (2017). The association between symptoms of autism and neuropsychological performance in females with Anorexia Nervosa. *Psychiatry Research*, *258*, 531 537. https://doi.org/10.1016/j.psychres.2017.09.005

Westwood, H., Stahl, D., Mandy, W., & Tchanturia, K. (2016). The set shifting profiles of anorexia nervosa and autism spectrum disorder using the Wisconsin card sorting test: A systematic review and meta analysis. *Psychological Medicine*, *46*(9), 1809 1827. https://doi.org/10.1017/S0033291716000581

Whiting, D. L., Deane, F. P., Ciarrochi, J., McLeod, H. J., & Simpson, G. K. (2015). Validating measures of psychological flexibility in a population with acquired brain injury. *Psychological Assessment*, *27*(2), 415 423. https://doi.org/10.1037/pas0000050

Wiebe, D. J., Baker, A. C., Suchy, Y., Stump, T. K., & Berg, C. A. (2018). Individual differences and day to day fluctuations in goal planning and type 1 diabetes management. *Health Psychology*, *37*(7), 638 646. https://doi.org/10.1037/hea0000624

Wildes, J. E., Forbes, E. E., & Marcus, M. D. (2014). Advancing research on cognitive flexibility in eating disorders: The importance of distinguishing attentional set shifting and reversal learning. *International Journal of Eating Disorders*, *47*(3), 227 230. https://doi.org/10.1002/eat.22243

Wilens, T. E., Carrellas, N. W., Martelon, M., Yule, A. M., Fried, R., Anselmo, R., & McCabe, S. E. (2017). Neuropsychological functioning in college students who misuse prescription stimulants. *American Journal on Addictions*, *26*(4), 379 387. https://doi.org/10.1111/ajad.12551

Yu, Y., Yu, Y., & Lin, Y. (2020). Anxiety and depression aggravate impulsiveness: The mediating and moderating role of cognitive flexibility. *Psychology Health and Medicine*, *25*(1), 25 36. https://doi.org/10.1080/13548506.2019.1601748

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2018). Cognitive underpinnings of nationalistic ideology in the context of Brexit. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), E4532 E4540. https://doi.org/10.1073/pnas.1708960115

Received September 29, 2021
Revision received November 30, 2021
Accepted January 10, 2022 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly

CHAPTER FIVE

Chapter 5 is published as:

[*]**Howlett, C.A.,** [*]Miles, S., Berryman, C., [#]Phillipou, A., & [#]Moseley, G.L. (2023). Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy. *Australian Journal of Psychology, 75*(1), 2174684. https://doi.org/10.1080/00049530.2023.2174684

* Co-first authors – contributed equally to this manuscript.

# Co-last authors – contributed equally to this manuscript.

This article is published in a fully open access journal under a Creative Commons Attribution-NonCommercial (CC-BY-NC) licence; therefore, no formal permission is required to reproduce it in this thesis. A signed statement of contribution for this co-authored manuscript can be found in Appendix F.



Australian Journal of Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/raup20

Conflation between self-report and neurocognitive assessments of cognitive flexibility: a critical review of the Jingle Fallacy

Caitlin A. Howlett, Stephanie Miles, Carolyn Berryman, Andrea Phillipou & G. Lorimer Moseley

To cite this article: Caitlin A. Howlett, Stephanie Miles, Carolyn Berryman, Andrea Phillipou & G. Lorimer Moseley (2023) Conflation between self-report and neurocognitive assessments of cognitive flexibility: a critical review of the Jingle Fallacy, Australian Journal of Psychology, 75:1, 2174684, DOI: 10.1080/00049530.2023.2174684

To link to this article: https://doi.org/10.1080/00049530.2023.2174684

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 19 Feb 2023.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=raup20

AUSTRALIAN JOURNAL OF PSYCHOLOGY
2023, VOL. 75, NO. 1, 2174684
https://doi.org/10.1080/00049530.2023.2174684

APS Australian Psychological Society
Taylor & Francis
Taylor & Francis Group

CRITICAL REVIEW ARTICLE

OPEN ACCESS Check for updates

Conflation between self-report and neurocognitive assessments of cognitive flexibility: a critical review of the Jingle Fallacy

Caitlin A. Howlett [a*], Stephanie Miles [b*], Carolyn Berryman [a,c], Andrea Phillipou [b,d,e,f#] and G. Lorimer Moseley [a#]

[a]Innovation, Implementation & Clinical Translation (IIMPACT) in Health, University of South Australia, Kaurna Country, Adelaide, Australia; [b]Centre for Mental Health, Swinburne University of Technology, Melbourne, Australia; [c]Brain Stimulation, Imaging and Cognition Research Group, School of Medicine, The University of Adelaide, Adelaide, Australia; [d]Department of Mental Health, St Vincent's Hospital, Melbourne, Australia; [e]Department of Psychiatry, The University of Melbourne, Melbourne, Australia; [f]Department of Mental Health, Austin Health, Melbourne, Australia

ABSTRACT

Cognitive flexibility is a widely studied construct and is considered an important treatment target for several psychological disorders. The convergence of several independent fields of research has led to assumptions about the assessment of cognitive flexibility – assumptions that are not empirically supported and often conflate different notions of flexibility. This critical review discusses how the conflation of self-report and neurocognitive assessments has seemingly arisen from literature on eating disorders. We describe how seminal early observations of "inflexible" personality characteristics, communication competence research, and investigations of frontal lobe function after injury led to two methods of assessing "cognitive flexibility". We discuss the impact that conflation of self-report and neurocognitive assessments has had on the field, and we provide recommendations for assessing cognitive flexibility in both research and clinical settings.

KEY POINTS

What is already known about this topic:

(1) Self-report and neurocognitive assessments of "cognitive flexibility" are commonly used in research and clinical practice.
(2) There is uncertainty in the field about whether or not self-report and neurocognitive assessments of "cognitive flexibility" assess similar underlying constructs.
(3) Both clinicians and researchers are susceptible to the *jingle fallacy*.

What this topic adds:

(1) This narrative critique of the literature reveals that self-report and neurocognitive assessments of "cognitive flexibility" have gradually been conflated over time.
(2) Early research in eating disorders seems to have played an influential role in generating and reinforcing such conflation.
(3) The assumption that self-report and neurocognitive assessments of "cognitive flexibility" are causally linked has no empirical basis and yet it has been used to explain inflexible cognitions and behaviours in people with eating disorders.

ARTICLE HISTORY
Received 20 June 2022
Accepted 25 January 2023

KEYWORDS
Executive function; neuropsychology; personality; feeding and eating disorders; surveys and questionnaires

Cognitive flexibility has been broadly conceptualised as the ability to modify cognitive and behavioural strategies in response to novel, changing, or unexpected demands (Cañas et al., 2003; Deák, 2003). The component thought to be central to cognitive flexibility is set-shifting – the ability to switch between different "mental sets" (Tchanturia, Anderluh, et al., 2004). Other components often attributed to cognitive flexibility include "the ability to change thoughts and behaviours" or "the ability to generate alternatives". Although the ability to be cognitively flexible relies on the intrinsic properties of the human cognitive system, whether cognitive flexibility is a separable component of executive function (Diamond, 2013), or an emergent property that derives from the interplay between different higher-order cognitive processes (Dajani & Uddin, 2015; Ionescu, 2012), is unresolved. Disagreements about the origins of cognitive flexibility likely stem from the lack of conceptual consensus that surrounds cognitive flexibility, as well as the multitude

CONTACT G. Lorimer Moseley lorimer.moseley@gmail.com
*Co-first authors.
#Co-last authors.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

of assessments that have been used to assess this construct (Ionescu, 2012).

Cognitive flexibility is commonly assessed by self-report questionnaires and/or performance on neurocognitive tasks. Although "cognitive flexibility" – a mental ability to switch thinking – has been differentiated from "behavioural flexibility" – a switch in behaviour when environmental contingencies change – available assessments do not seem to make such a differentiation. Perhaps this reflects the view that cognitive and behavioural flexibility are so intertwined that they might be considered inseparable (Uddin, 2021). Common self-report assessments that target "the ability to change thoughts and behaviours" or "the ability to generate alternatives", include the Cognitive Flexibility Scale (Martin & Rubin, 1995) and the Cognitive Flexibility Inventory (Dennis & Vander Wal, 2010). Common neurocognitive tasks that assess a switch in behaviour when environmental contingencies change, include the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948; Heaton et al., 1993), the Trail Making Test (TMT; Reitan, 1955, 1958), and the Intra-Extra Dimensional Set Shift task (Robbins et al., 1998).

Despite empirical evidence that self-report and neurocognitive assessments do not relate well in neither clinical nor non-clinical, samples (e.g., Howlett et al., 2021, 2022; Johnco et al., 2014; Lounes et al., 2011; Miles et al., 2022; Sternheim et al., 2022), these assessments are often used interchangeably, and considered to be equivalent in assessing the construct of cognitive flexibility. Here, we provide a historical overview of how and why self-report and neurocognitive assessments that purport to assess cognitive flexibility may have become conflated over time. We will focus our discussion largely on eating disorders research, which appears to be the key field to have conflated these two assessment approaches, The conflation has since extended to other fields. We outline the unintended consequences of the conflation and we make recommendations for how future research and clinical practice might best respond.

The historical development of cognitive flexibility assessments

Early 20th century researchers observed that those with frontal lobe damage performed worse than healthy controls on simple sorting tasks (Weigl, 1927). In a series of studies with rhesus monkeys, Settlage et al. (1948) observed that monkeys with frontal cortex lesions lost the ability to complete shifting problems even though they could solve non-shifting problems. On the basis of that work, Berg (1948) developed the WCST to assess "flexibility in thinking" in humans. The WCST is now widely accepted as a test of "cognitive flexibility" (Dennis & Vander Wal, 2010; Johnco et al., 2014; Tchanturia et al., 2012). Similarly, Reitan (1955, 1958) initially developed the TMT to detect brain damage, but since then it has been used to detect poor "cognitive flexibility" (Christidi et al., 2015; Dennis & Vander Wal, 2010; Johnco et al., 2014). Other neurocognitive tasks that have become popularised as cognitive flexibility assessments include the Haptic Illusion Task (Tchanturia et al., 2001; Uznadze, 1966), the Brixton Spatial Anticipation Task (Burgess & Shallice, 1997; Holliday et al., 2005; Lounes et al., 2011; Tchanturia et al., 2002; Tchanturia, Anderluh, et al., 2004), and the Alternate Uses Task (Eslinger & Gratten, 1993; Guilford, 1956).

The development of self-report questionnaires of cognitive flexibility followed a parallel, but separate trajectory. This trajectory seems to have commenced in interpersonal communication competence research. Specifically, Bochner and Kelly (1974) developed a conceptual framework, which proposed that competent communicators possess a set of core skills that are directly observable by others – one of them being the ability to display behavioural flexibility. Wiemann (1977) empirically investigated the relationship between what was called interaction management – a set of rules that guide social interactions (e.g., one person speaks at a time) – and aspects of communication competence, including behavioural flexibility, and found a positive correlation. With these ideas and findings in mind, Martin and Rubin (1994) went on to devise the Communication Flexibility Scale – a questionnaire that assesses a person's ability to change their communication behaviour in various contextual situations. This study found that the Communication Flexibility Scale positively correlated with other self-report questionnaires designed to capture social desirability and communication adaptability: and that those who scored high on communication flexibility also scored high on social desirability and communication adaptability (Martin & Rubin, 1994). This line of research, along with the fact that cognitive flexibility was thought to be a prerequisite for being behaviourally flexible (Martin & Rubin, 1995), ultimately led to the development, and later the initial validation, of the Cognitive Flexibility Scale (Martin & Anderson, 1998; Martin & Rubin, 1995). Since then, other self-report questionnaires designed to assess cognitive and behavioural flexibility have been developed, including the Cognitive Flexibility Inventory (Dennis & Vander Wal, 2010), the shift subscale of the

Behavior Rating Inventory of Executive Function (Roth et al., 2005) and the Detail and Flexibility Questionnaire (DFlex; Roberts et al., 2011). Each of these questionnaires require the participant to reflect on their own thinking and behaviour before making a judgement about how flexible they view themselves to be. One limitation of self-report questionnaires is that they are inherently subjective; thus, level of insight, metacognitive awareness and response biases may influence results. For example, poor insight has been reported to be a feature of some patients with eating disorders, and it has been suggested that reduced neurocognitive flexibility may contribute to poor insight (Konstantakopoulos et al., 2011). However, it seems equally possible that poor insight may lead certain patients to overestimate their own flexibility in thinking and behaviour, and thus show inflated scores on a self-report questionnaire. Alternatively, someone who is highly perfectionistic or lacks self-esteem may judge their own cognitive flexibility too harshly, and thus show deflated scores on a self-report questionnaire.

Intuition trumps evidence – conflation of different constructs of cognitive flexibility

The coming together of these separate lines of research appears to have largely occurred in the field of eating disorders, then spread into other fields. In the late 20th century, Strober (1980) explored personality types using the California Psychological Inventory (Gough, 1956), a 480-item true/false self-report questionnaire that produces scores for 18 personality traits. One such subscale was the Flexibility (Fx) subscale, which comprised of 28-items that were designed to assess the ability to be adaptable in thinking, behaviour, and temperament (Donnay & Elliott, 2003). In Strober's (1980) study, people with anorexia nervosa scored lower than clinical control cohorts on this "Flexibility" subscale, leading Strober to conclude that there is a distinct type of "personality" that characterises people with anorexia nervosa. That finding was corroborated and extended by Casper et al. (1992) who reported that people with anorexia nervosa scored lower on the California Psychological Inventory Flexibility subscale than both healthy controls and those with bulimia nervosa. This apparent "inflexible personality trait" in people with eating disorders seems to have led researchers to pursue more quantifiable assessments. Here they turned to the neurocognitive tasks outlined above, presumably because they purported to assess "flexibility in thinking".

From the early 1990's onwards, there was a steady increase in "cognitive flexibility" research, primarily focused on neurocognitive tasks (e.g., Anderluh et al., 2003; Cavedini et al., 2004; Fassino et al., 2002; Szmukler et al., 1992; Tchanturia et al., 2001, 2002; Tchanturia, Anderluh, et al., 2004). During the last two decades, many studies found poorer performance on neurocognitive tasks of cognitive flexibility in people with anorexia nervosa than in healthy controls (Fassino et al., 2002; Holliday et al., 2005; Lounes et al., 2011; Steinglass et al., 2006; Tchanturia et al., 2001, 2002, 2012; Tchanturia, Anderluh, et al., 2004; Westwood et al. 2016), although many others did not (Cavedini et al., 2004; Dmitrzak-Węglarz et al., 2011; Galimberti et al., 2012; Giel et al., 2012; Miles et al., 2020; Tokley & Kemps, 2007; Vall & Wade, 2015; Van Autreve et al., 2016). Poor performance on neurocognitive tasks was often proposed as a potential contributor to clinical observations of inflexible thoughts and behaviours (i.e., calorie counting and strict exercise regimes) in people with anorexia nervosa (Brockmeyer et al., 2014; Rößner et al., 2017; Tchanturia et al., 2012, 2013). There appears to be no empirical data to support that assertion. Nonetheless, the results of neurocognitive tasks certainly *appeared* to corroborate findings from self-report questionnaires that consistently showed poorer self-evaluated cognitive flexibility in people with eating disorders than in healthy controls (Dell'Osso et al., 2018; Herbrich et al., 2018; Lao-Kaim et al., 2015; Lounes et al., 2011; McAnarney et al., 2011; Roberts et al., 2011).

The extant literature on cognitive flexibility does not explicitly suggest that self-reported cognitive flexibility is equivalent to performance on neurocognitive cognitive flexibility tasks. However, it is evident, particularly within the eating disorders field, that these assessments of cognitive flexibility have become conflated over time (e.g., Abbate Daga et al., 2014; Friederich & Herzog, 2011; Genders & Tchanturia, 2010; Holliday et al., 2005; Leppanen et al., 2018; Perpiñá et al., 2017; Steinglass et al., 2006; Tchanturia et al., 2012; Tchanturia, Anderluh, et al., 2004). Such conflation becomes particularly clear when we examine the rationales provided in the introductions of various papers. For example, Holliday et al. (2005) writes: "The hypothesized association between set-shifting difficulties and anorexia nervosa have face validity in that individuals with anorexia nervosa are often described as persistent, with rigid, conforming, or obsessional personalities" (p. 1). Leppanen et al. (2018) echoes a similar argument: "As changes in thinking styles are believed to contribute to treatment resistance in AN, interventions targeting neurocognitive

processes are needed" (p. 1). That those with eating disorders are often characterised by both types of assessments as exhibiting poorer cognitive flexibility than healthy controls, and that the two assessment types are defined by the same term, intuitively points to a unified underlying construct. At face value, self-report and neurocognitive assessments certainly *seem* to relate. For example, if a group of people are displaying difficulties on average in both of these types of assessments, which share the one label of "flexibility", then it would *seem* reasonable to assume that both types of assessments are revealing a broader problem with cognitive flexibility. However, such assumptions leave researchers and clinicians vulnerable to what has been referred to as the *jingle fallacy* – the erroneous assumption that two tests assess the same construct because they share similar names or labels – which was first described by a "Professor Aikins" cited by Thorndike (1904, p. 14) over a century ago. The interchangeable use of these assessments (e.g., Abbate Daga et al., 2011; Arlt et al., 2016; Dell'Osso et al., 2018; Fujino et al., 2019; Malivoire et al., 2019; Oguz et al., 2019; Tchanturia et al., 2012) is predicated on that implicit assumption, the truth of which remains unwarranted. Further, given the abundance of tasks that are currently being used to capture cognitive flexibility, it can also be argued that such conflation is not confined solely to self-report and neurocognitive assessments, but also exists to a large degree within the neurocognitive space itself. Failure to detect the jingle fallacy in psychological research not only has the potential to cause immense confusion and heterogeneity in regards to defining, conceptualising and assessing psychological phenomena – including, but by no means limited to, the construct of cognitive flexibility – but also has putative consequences for the field, such that "we could end up with several constructs under the same name" (Ionescu, 2012, p. 190). Such threats to our field can begin to be mitigated by i) transparently reporting construct measurement (e.g., What is the construct? What measure was used to operationalise the construct? How is the construct quantified?) to avoid falling into a "measurement schmeasurement attitude" (Flake & Fried, 2020) or engaging in questionable measurement practices, and ii) better understanding the ways in which self-report and neurocognitive assessments of "cognitive flexibility" are similar and distinct. Implementing the ten conceptual, empirical, and developmental criteria proposed by Lawson and Robins (2021) would help to elucidate whether these two assessment approaches can be considered as *sibling constructs* under the *parent construct* of "cognitive flexibility".

While self-report and neurocognitive assessments have been implied to target the single construct of "cognitive flexibility", others have suggested they are causally linked. For example, one proposal is that inflexible behaviours and thoughts (typically assessed via self-report questionnaires, interviews, or informant reports) are underpinned by processing impairments that can be assessed by neurocognitive tasks bearing the same name – assessments of "cognitive flexibility" (Abbate Daga et al., 2014; D'cruz et al., 2013; Leppanen et al., 2018; Steinglass et al., 2006; Tchanturia et al., 2012). To illustrate this point, Abbate Daga et al. (2014) state that "several lines of evidence suggest that some clinical aspects of AN [anorexia nervosa] could mirror alterations of cognitive functions; in particular, those rigid and perfectionistic features that usually characterize affected individuals could be the result of set-shifting inefficiencies" (p. 592). Similarly, Tchanturia et al. (2012) argued that "People with eating disorders (ED) frequently present with inflexible behaviours, including eating related issues which contribute to the maintenance of the illness. Small scale studies point to difficulties with cognitive set-shifting as a basis" (p.1). By contrast, the opposite – that poor performance on neurocognitive tests of cognitive flexibility in people with anorexia nervosa results from their inflexible behaviours and thinking – has also been proposed. For example, " … clinical observations of impaired cognitive-behavioural flexibility in AN [anorexia nervosa] patients underlie impaired set-shifting and impaired behavioural response shifting" (Friederich & Herzog, 2011, p. 115). However, there is no clear empirical evidence to support either of these opposing theories. Thus, there appears to be three apparent conceptual manifestations in the field: that inflexible thoughts and behaviours cause neurocognitive inflexibility, that neurocognitive inflexibility causes inflexible thoughts and behaviours, and that a single construct – "cognitive flexibility" –exists, which can be assessed through self-reported thoughts and behaviours or neurocognitive tasks. However, it should be noted that such explanations likely oversimplify, and thus downplay the influential role of other factors on cognitive flexibility. Future research might consider conducting mediation analyses to establish whether cognitive (e.g., self-efficacy) and non-cognitive (e.g., sleep quality) factors predispose particular manifestations of cognitive flexibility.

The evidence concerning how well the two types of assessments relate to one another conflicts with all three conceptual manifestations. Primary studies have reported little-to-no relationship between self-report and neurocognitive assessments of cognitive

flexibility. For example, Gelonch et al. (2016) found no relationship between the shift subscale of the Behavior Rating Inventory of Executive Function – Adult version and outcomes on the TMT and WCST in healthy individuals and people with fibromyalgia. Lounes et al. (2011) did not detect a relationship between the Cognitive Flexibility Scale and the Brixton Spatial Anticipation Task in people with anorexia nervosa and a healthy cohort. In university students, Gonzalez et al. (2013) reported no association between the Cognitive Flexibility Scale and the TMT and a novel flexibility task that required participants to switch between rules in order to complete a maze-like puzzle. Johnco et al. (2014) reported that the Cognitive Flexibility Scale and Cognitive Flexibility Inventory did not correlate with a range of neurocognitive tasks in older adults with comorbid depressive and anxiety symptoms, but small correlations ($r \approx 0.28$) were detected in the healthy comparison group. Two recent extensive meta-analytic reviews (Howlett et al., 2021, 2022), drew data from published and unpublished work and also failed to detect a relationship in non-clinical or clinical cohorts. Two recent studies reported null correlations between a range of different self-report and neurocognitive assessments of cognitive flexibility. Miles et al. (2022) found no significant correlations between the Cognitive Flexibility Inventory and two neurocognitive tasks – the TMT and WCST – in people with and without a lifetime diagnosis of anorexia nervosa; Sternheim et al. (2022) reported no significant association between the TMT and the DFlex in two patient groups: people with anorexia nervosa and those with obsessive compulsive disorder. Together, these findings cast significant doubt about whether these assessments capture the same construct.

A possible explanation as to why neurocognitive assessements and self-report questionnaires of cognitive flexibility do not relate is because neurocognitive tasks lack ecological validity (Chaytor et al., 2006), and by design they are suited to assess momentary performance alone. In contrast, self-report questionnaires by design, assess relatively stable thoughts and behaviours of respondents over longer timescales (such issues have been previously discussed in relation to the construct of self-control (i.e., Wennerhold & Friese, 2020)). It can therefore be argued that if self-report questionnaires were re-designed to capture flexibility *"in the moment"*, rather than capturing responses that represent *"typical thoughts and behaviours"*, then perhaps these two assessment approaches would demonstrate desirable correlations. Although self-report questionnaires of cognitive flexibility often have high reliability (Dennis & Vander Wal, 2010; Johnco et al., 2014; Martin & Anderson, 1998), neurocognitive tasks do not. The lack of reliability that is associated with neurocognitive tasks has also previously been described as a possible contributor to the poor correlations that are typically observed between self-report and neurocognitive assessments (Dang et al., 2020). Further, the lack of association between self-report and neurocognitive tasks of "cognitive flexibility" may be a consequence of the vast variety of neurocognitive tasks that are currently available. That is, each test may tap into slightly different underlying processes, which presents unique challenges when attempting to accurately match these tasks to self-report assessments. Another explanation for the poor relationship between self-report and neurocogntive assessments of cognitive flexibility is that these assessments may be tapping into two entirely distinct constructs. One may begin to ponder then, an important question: how is it that a group of people such as those with anorexia nervosa, for example, can perform worse on both assessments, yet performance on one assessment does not relate to performance on another? One might suggest that people can perform differentially on these two assessments (i.e., perform well on one assessment but not the other), but this speculation goes against empirical data that have established that some clinical conditions, such as anorexia nervosa, are associated with deficits on both assessments. However, it is not yet possible to determine whether one (or neither) assessment is accurate in capturing the construct of cognitive flexibility. Notwithstanding, conflating these two assessment tools by generalising the results of one to the other could limit our conceptualisation and understanding of the construct of cognitive flexibility. Further, conflating these assessments could result in mischaracterisations of clinical disorders.

As stated earlier, the problem of conflating self-report and neurocognitive assessments is not unique to the field of cognitive flexibility. The executive functioning literature more broadly has encountered similar problems where self-report questionnaires are presumed to capture identical or similar underlying processes that are assessed by neurocognitive tasks (Toplak et al., 2013). Those presumptions are often unsupported by empirical data obtained from adults (Burmester et al., 2016; Rabin et al., 2006; Spencer et al., 2010), adolescents (Herbrich et al., 2018; Stedal & Dahlgren, 2015) and children (Soto et al., 2020). Further, some clinicians overestimate the association between self-report questionnaires and neurocognitive performance on tasks that assess memory, attention, and thinking/organisation (Spencer et al., 2010),

which may also explain why assessments of cognitive flexibility are used interchangeably in clinical practice. Anecdotally, knowing that self-report and neurocognitive assessments do not *usually* relate well does not necessarily prevent their interchangeable use, nor the generalisation of findings from one assessment tool to another. We propose therefore, two significant implications for clinical practice. First, clinicians could integrate informant-reports as a part of a comprehensive clinical assessment to capture a client's mental ability to switch thinking and adapt behaviours in accordance with changing contexts. Clinicians may want to implement informant-reports as complementary tools to clinical assessment because these tools are considered to be less subjective and less influenced by other decrements (e.g., a lack of insight). Tools that have been widely implemented and can assist clinicians in avoiding erroneous assumptions include the informant version of the Behavior Rating Inventory of Executive Function (Roth et al., 2005). Second, there is some time urgency about setting the record straight with regard to "cognitive flexibility" and moving forward unencumbered by erroneous assumptions of equivalence between what seem to be distinct constructs.

Limitations of existing research and future recommendations

The conflation of self-report and neurocognitive assessments of cognitive flexibility present problems that can be immediately addressed. First, we should acknowledge that available evidence suggests against a causal relationship between the neurophysiological processes that enable one to perform well on neurocognitive tests, and one's tendency to report flexibility in one's thoughts and behaviours. Furthermore, given that there is no evidence of a 1:1 relationship between flexible thinking and flexible behaviour, we should also be precise in our language when describing these distinct, but potentially related facets of flexibility. Second, we should cease interchanging self-report and neurocognitive assessments of cognitive flexibility (exemplar studies include Miles et al., 2022; Sternheim et al., 2022). Third, we can recognise as erroneous the conclusion that people with clinically observable or self-reported "inflexible" behaviours can be accurately and precisely evaluated with neurocognitive tasks. This would avoid the current potential of dismissing or overlooking the clinical relevance of self-report questionnaires (which can provide information that is more ecologically valid and client-specific), and should thus promote a more client centric clinical encounter. To provide a more comprehensive assessment of cognitive flexibility, it may be appropriate for clinicians and researchers to use self-report and neurocognitive assessments alongside more ecologically valid tasks – e.g., the Virtual Cooking Task (Chicchi Giglioli et al., 2021) – that allow individuals to be immersed in virtual environments where they are required to perform everyday activities. Poor performance on one of these assessments may indicate underlying issues that may warrant further investigation. For example, neurocognitive tasks may be more sensitive to capturing breakdowns in cognitive flexibility, whereas virtual reality technology may be able to assist in uncovering instances of cognitive inflexibility in real-life scenarios, and self-report questionnaires might provide more nuanced differences in cognitive flexibility. However, users must remain aware that self-report and neurocognitive tasks may be targeting different contructs altogether.

Other unknowns will require more concerted attention. That the two approaches to assessing cognitive flexibility do not appear to relate well may reflect problems with the assessments themselves rather than divergence in the constructs that they aim assess. For example, many neurocognitive tasks may be restricted by the *task impurity problem* – the phenomenon that a task does not assess one cognitive domain or skill in isolation, but rather relies on a combination of both executive and non-executive processes, making the cause of poor task performance difficult to pinpoint (Miyake et al., 2000; Müller & Kerns, 2015). It should be noted, however, that attempting to "isolate" cognitive flexibility may distort the very construct itself, because cognitive flexibility may emerge from its reliance on various cognitive processes (Dajani & Uddin, 2015). Additionally, breakdowns in cognitive flexibility, as indexed by neurocognitive tasks, may be attributable to deficits in other executive processes such as working memory or inhibition. For instance, problems with inhibition may differentially impact cognitive flexibility outcomes and have no impact on working memory. It is not yet possible to untangle the contributors to poor cognitive flexibility performance, but ongoing identification of the genetic and neural substrates that underpin cognitive flexibility performance may shed some light (Nomi et al., 2017; Verdejo-Garcia et al., 2015; Zhang et al., 2019). Therefore, it is vital that researchers and clinicians are aware of such limitations of neurocognitive tasks when interpreting and drawing conclusions from findings and control for other executive processes where possible.

One criticism of self-report questionnaires is that they can be confounded by negative affect, with studies reporting an association between self-reported cognitive flexibility, and depression and anxiety (Dennis & Vander

Wal, 2010; Johnco et al., 2014; Lounes et al., 2011; Miles et al., 2022; Wang et al., 2019; Yu et al., 2020). Although some studies have found that neurocognitive tasks do not appear to be confounded by negative affect (Herbrich et al., 2018; Lang et al., 2015; Lounes et al., 2011; Miles et al., 2022) others studies have found that neurocognitive tasks are modulated by emotional valence (Wang et al., 2017). These assessments likely still have great utility for our field, but rigorous validation and replication studies are still required if we are to understand the possible impact of such factors on results from either assessment approach, let alone if we are to delineate the constructs that self-report questionnaires and neurocognitive tasks actually assess.

Our final recommendation is pragmatic. As clinicians and researchers, we need to clearly identify and articulate what it is that we are interested in assessing. Are we interested in self-reported evaluations of adaptability in thoughts and/or behaviours, or performance on neurocognitive tasks that require shifting strategies, or both? By being more precise about the construct in which we are interested, we can more easily select the best assessment with which to investigate it and avoid erroneous generalisations beyond our data. The broad term of "cognitive flexibility" may well be an umbrella term that is simply too imprecise. Making precise interpretations and not generalising beyond the assessment modality used will facilitate progress in our understanding of these constructs and how they relate to psychopathology.

Conclusion

The conflation of neurocognitive tasks and self-reported outcomes of cognitive flexibility has arisen through the convergence of seemingly similar lines of research and intuitive, but not empirically supported, assumptions. There remains scant evidence in support of the claim that poor performance on neurocognitive tasks will be mirrored by poor performance on self-report questionnaires within individuals, even though there is building evidence that certain clinical cohorts perform poorly as a group on both types of assessments. Future research and clinical practice should give careful consideration when selecting which cognitive flexibility assessments to use by ensuring that the chosen assessments are aligned with the aims of research or the goals of therapy. It is strongly encouraged that both researchers and clinicians be precise when describing or referring to such assessments, and when interpreting results. Lastly, the interchangeable use, and conflation of, neurocognitive tasks and self-reported assessments of "cognitive flexibility" should be avoided.

Disclosure statement

GLM has received support from: Reality Health, ConnectHealth UK, Seqirus, Kaiser Permanente, Workers' Compensation Boards in Australia, Europe and North America, AIA Australia, the International Olympic Committee, Port Adelaide Football Club, and Arsenal Football Club. Professional and scientific bodies have reimbursed him for travel costs related to presentation of research at scientific conferences/symposia. He has received speaker fees for lectures on pain and rehabilitation. He receives book royalties from NOIgroup publications, Dancing Giraffe Press & OPTP. CB has received support from Workers' Compensation Boards in Australia. All remaining authors (CAH, SM and AP) declare no potential conflicts of interest (personal or financial) with respect to the contents, authorship, or publication of this manuscript.

Funding

AP was supported by a National Health and Medical Research Council (NHMRC) Project Grant [ID 1159953]. CB and GLM were supported by a NHMRC Leadership Investigator Grant [ID 1178444] to GLM. These funding bodies had no role in the writing of the manuscript, or the decision to submit the paper for publication.

ORCID

Caitlin A. Howlett http://orcid.org/0000-0002-4584-8641
Stephanie Miles http://orcid.org/0000-0001-8190-894X
Carolyn Berryman http://orcid.org/0000-0002-5316-0847
Andrea Phillipou http://orcid.org/0000-0003-1009-6619
G. Lorimer Moseley http://orcid.org/0000-0002-3750-4945

Data availability statement

Data sharing is not applicable to this article as no new data were collected or analysed in this manuscript.

References

Abbate Daga, G., Buzzichelli, S., Amianto, F., Rocca, G., Marzola, E., McClintock, S. M., & Fassino, S. (2011). Cognitive flexibility in verbal and nonverbal domains and decision making in anorexia nervosa patients: A pilot study. *BMC Psychiatry*, *11*(162), 1–8. https://doi.org/10.1186/1471-244X-11-162

Abbate Daga, G., Buzzichelli, S., Marzola, E., Amianto, F., & Fassino, S. (2014). Clinical investigation of set-shifting subtypes in anorexia nervosa. *Psychiatry Research*, *219*(3), 592–597. https://doi.org/10.1016/j.psychres 2014.06.024

Anderluh, M. B., Tchanturia, K., Rabe-Hesketh, S., & Treasure, J. (2003). Childhood obsessive-compulsive personality traits in adult women with eating disorders: Defining a broader eating disorder phenotype. *American Journal of Psychiatry*, *160*(2), 242–247. https://doi.org/10.1176/appi.ajp.160 2.242

Arlt, J., Yiu, A., Eneva, K., Taylor Dryman, M., Heimberg, R. G., & Chen, E. Y. (2016). Contributions of cognitive inflexibility to eating disorder and social anxiety symptoms. *Eating Behaviors*, *21*, 30–32. https://doi.org/10.1016/j.eatbeh.2015.12.008

Berg, E. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, *39* (1), 15–22. https://doi.org/10.1080/00221309.1948.9918159

Bochner, A. P., & Kelly, C. W. (1974). Interpersonal communication instruction-theory and practice – a symposium. 1. Interpersonal competence-rationale, philosophy, and implementation of a conceptual framework. *Speech Teacher*, *23*(4), 279–301. https://doi.org/10.1080/03634527409378103

Brockmeyer, T., Ingenerf, K., Walther, S., Wild, B., Hartmann, M., Herzog, W., Bents, H., & Friederich, H. C. (2014). Training cognitive flexibility in patients with anorexia nervosa: A pilot randomized controlled trial of cognitive remediation therapy. *International Journal of Eating Disorders*, *47*(1), 24–31. https://doi.org/10.1002/eat.22206

Burgess, P. W., & Shallice, T. (1997). The Hayling and Brixton tests

Burmester, B., Leathem, J., & Merrick, P. (2016). Subjective cognitive complaints and objective cognitive function in aging: A systematic review and meta-analysis of recent cross-sectional findings. *Neuropsychological Review*, *26*(4), 376–393. https://doi.org/10.1007/s11065-016-9332-2

Cañas, J. J., Quesada, J. F., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics*, *46*(5), 482–501. https://doi.org/10.1080/0014013031000061640

Casper, R. C., Hedeker, D., & McClough, J. F. (1992). Personality dimensions in eating disorders and their relevance for subtyping. *Journal of the American Academy of Child & Adolescent Psychiatry*, *31*(5), 830–840. https://doi.org/10.1097/00004583-199209000-00008

Cavedini, P., Bassi, T., Ubbiali, A., Casolari, A., Giordani, S., Zorzi, C., & Bellodi, L. (2004). Neuropsychological investigation of decision-making in anorexia nervosa. *Psychiatry Research*, *127*(3), 259–266. https://doi.org/10.1016/j.psychres.2004.03.012

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, *21*(3), 217–227. https://doi.org/10.1016/j.acn.2005.12.002

Chicchi Giglioli, I. A., Galvez, B. P., Granados, A. G., & Raya, M. A. (2021). The virtual cooking task: A preliminary comparison between neuropsychological and ecological virtual reality tests to assess executive functions alterations in patients affected by alcohol use disorder. *Cyberpsychology, Behavor, and Social Networking*, *24*(10), 673–682. https://doi.org/10.1089/cyber.2020.0560

Christidi, F., Kararizou, E., Triantafyllou, N., Anagnostouli, M., & Zalonis, I. (2015). Derived trail making test indices: Demographics and cognitive background variables across the adult life span. *Aging, Neuropsychology, and Cognition*, *22*(6), 667–678. https://doi.org/10.1080/13825585.2015.1027650

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in neurosciences*, *38*(9), 571–578. https://doi.org/10.1016/j.tins.2015.07.003

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioural measures weakly correlated? *Trends in Cognitive Sciences*, *24*(4), 267–269. https://doi.org/10.1016/j.tics.2020.01.007

D'cruz, A., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, *27*(2), 152–160. https://doi.org/10.1037/a0031721

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior*, *31*, 271–327. https://doi.org/10.1016/S0065-2407(03)31007-9

Dell'Osso, L., Carpita, B., Gesi, C., Cremone, I. M., Corsi, M., Massimetti, E., Muti, D., Calderani, E., Castellini, G., Luciano, M., Ricca, V., Carmassi, C., & Maj, M. (2018). Subthreshold autism spectrum disorder in patients with eating disorders. *Comprehensive Psychiatry*, *81*, 66–72. https://doi.org/10.1016/j.comppsych 2017.11.007

Dennis, J. P., & Vander Wal, J. S. (2010). The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research*, *34*(3), 241–253. https://doi.org/10.1007/s10608-009-9276-4

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dmitrzak-Węglarz, M., Słopień, A., Tyszkiewicz, M., Rybakowski, F., Rajewski, A., & Hauser, J. (2011). Polymorphisms of the SNAP-25 gene and performance on the Wisconsin Card Sorting Test in anorexia nervosa and in healthy adolescent participants. *Archives of Psychiatry and Psychotherapy*, *1*, 43–51. https://doi.org/10.1016/S0924-977X(11)70526-0

Donnay, D. A. C., & Elliott, T. R. (2003). The California psychological inventory. In L. E. Beutler (Ed.), *Integrative assessment of adult personality* (pp. 227–261). Guilford Press.

Eslinger, P. J., & Gratten, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia*, *31*(1), 17–28. https://doi.org/10.1016/0028-3932(93)90077-D

Fassino, S., Pier, A., Daga, G. A., Leombruni, P., Mortara, P., & Rovera, G. G. (2002). Attentional biases and frontal functioning in anorexia nervosa. *International Journal of Eating Disorders*, *31*(3), 274–283. https://doi.org/10.1002/eat.10028

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Friederich, H. C., & Herzog, W. (2011). Cognitive-behavioral flexibility in anorexia nervosa. In R. A. H. Adan & W. H. Kaye (Eds.), *Behavioural neurobiology of eating disorders* (pp. 111–123). Springer. https://doi.org/10.1007/978-3-642-15131-6

Fujino, J., Tei, S., Itahashi, T., Aoki, Y., Ohta, H., Kubota, M., Isobe, M., Hashimoto, R., Nakamura, M., Kato, N., & Takahashi, H. (2019). Need for closure and cognitive flexibility in individuals with autism spectrum disorder:

A preliminary study. *Psychiatry Research*, *271*, 247–252. https://doi.org/10.1016/j.psychres.2018.11.057

Galimberti, E., Martoni, R. M., Cavallini, M. C., Erzegovesi, S., & Bellodi, L. (2012). Motor inhibition and cognitive flexibility in eating disorder subtypes. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *36*(2), 307–312. https://doi.org/10.1016/j.pnpbp.2011.10.017

Gelonch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and objective measures. *Comprehensive Psychiatry*, *66*, 113–122. https://doi.org/10.1016/j.comppsych.2016.01.002

Genders, R., & Tchanturia, K. (2010). Cognitive Remediation Therapy (CRT) for anorexia in group format: A pilot study. *Eating and Weight Disorders*, *15*(4), e234–239. https://doi.org/10.1007/BF03325304

Giel, K. E., Wittorf, A., Wolkenstein, L., Klingberg, S., Drimmer, E., Sch¨onenberg, M., Rapp, A. M., Fallgatter, A. J., Hautzinger, M., & Zipfel, S. (2012). Is impaired set-shifting a feature of '"pure"' anorexia nervosa? Investigating the role of depression in set-shifting ability in anorexia nervosa and unipolar depression. *Psychiatry Research*, *200*(2–3), 538–543. https://doi.org/10.1016/j.psychres 2012.06.004

Gonzalez, C. A., Figueroa, I. J., Bellows, B. G., Rhodes, D., & Youmans, R. J. (2013). A new behavioral measure of cognitive flexibility. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 297–306). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39360-0-33

Gough, H. G. (1956). *California psychological inventory*. Consulting Psychologists Press.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, *53*(4), 267–293. https://doi.org/10.1037/h0040755

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Psychological Assessment Resources.

Herbrich, L., Kappel, V., van Noort, B. M., & Winter, S. (2018). Differences in set-shifting and central coherence across anorexia nervosa subtypes in children and adolescents. *European Eating Disorders Review*, *26*(5), 499–507. https://doi.org/10.1002/erv.2605

Holliday, J., Tchanturia, K., Landau, S., Collier, D., & Treasure, J. (2005). Is impaired set-shifting an endophenotype of anorexia nervosa? *The American Journal of Psychiatry*, *162*(12), 2269–2275. https://doi.org/10.1176/appi.ajp.162.12.2269

Howlett, C. A., Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., Karran, E. L., Szeto, K., Pronk, L., Miles, S., & Moseley, G. L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, *88*(102061), 1–21. https://doi.org/10.1016/j.cpr.2021.102061

Howlett, C. A., Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., Karran, E. L., Szeto, K., Pronk, L., Miles, S., & Moseley, G. L. (2022). Back to the drawing board —the relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology*, *36*(5), 347–372. https://doi.org/10.1037/neu0000796

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psycholog*, *30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). Reliability and validity of two self-report measures of cognitive flexibility. *Psychological Assessment*, *26*(4), 1381–1387. https://doi.org/10.1037/a0038009

Konstantakopoulos, G., Tchanturia, K., Surguladze, S. A., & David, A. S. (2011). Insight in eating disorders: Clinical and cognitive correlates. *Psychological Medicine*, *41*(9), 1951–1961. https://doi.org/10.1017/S0033291710002539

Lang, K., Lloyd, S., Khondoker, M., Simic, M., Treasure, J., & Tchanturia, K. (2015). Do children and adolescents with anorexia nervosa display an inefficient cognitive processing style? *Plos One*, *10*(7), 1–13. https://doi.org/10.1371/journal.pone.0131724

Lao-Kaim, N. P., Fonville, L., Giampietro, V. P., Williams, S. C. R., Simmons, A., & Tchanturia, K. (2015). Aberrant function of learning and cognitive control networks underlie inefficient cognitive flexibility in anorexia nervosa: A cross-sectional fMRI study. *Plos One*, *10*(5), e0124027. https://doi.org/10.1371/journal.pone.0124027

Lawson, K. M., & Robins, R. W. (2021). Sibling constructs: What are they, why do they matter, and how should you handle them? *Personality and Social Psychology Review*, *25*(4), 344–366. https://doi.org/10.1177/10888683211047101

Leppanen, J., Adamson, J., & Tchanturia, K. (2018). Impact of cognitive remediation therapy on neurocognitive processing in anorexia nervosa. *Frontiers in Psychiatry*, *9*(96), 1–5. https://doi.org/10 3389/fpsyt.2018.00096

Lounes, N., Khan, G., & Tchanturia, K. (2011). Assessment of cognitive flexibility in anorexia nervosa-self-report or experimental measure? A brief report. *Journal of the International Neuropsychological Society*, *17*(5), 925–928. https://doi.org/10.1017/S1355617711000671

Malivoire, B. L., Stewart, K. E., Tallon, K., Ovanessian, M. M., Pawluk, E. J., & Koerner, N. (2019). Negative urgency and generalized anxiety disorder symptom severity: The role of self-reported cognitive processes. *Personality and Individual Differences*, *145*, 58–63. https://doi.org/10.1016/j.paid.2019.03.021

Martin, M. M., & Anderson, C. M. (1998). The cognitive flexibility scale: Three validity studies. *International Journal of Phytoremediation*, *11*(1), 1–9. https://doi.org/10.1080/08934219809367680

Martin, M. M., & Rubin, R. B. (1994). Development of a communication flexibility measure. *The Southern Communication Journal*, *59*(2), 171–178. https://doi.org/10.1080/10417949409372934

Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports*, *76*(2), 623–626. https://doi.org/10 2466/pr0.1995.76.2.623

McAnarney, E. R., Zarcone, J., Singh, P., Michels, J., Welsh, S., Litteer, T., Wang, H., & Klein, J. D. (2011). Restrictive anorexia nervosa and set-shifting in adolescents: A biobehavioral interface. *Journal of Adolescent Health*, *49*(1), 99–101. https://doi.org/10.1016/j.jadohealth.2010.11.259

Miles, S., Gnatt, I., Phillipou, A., & Nedeljkovic, M. (2020). Cognitive flexibility in acute anorexia nervosa and after recovery: A systematic review. *Clinical Psychology Review*, *81*(101905), 1–129. https://doi.org/10.1016/j.cpr.2020.101905

Miles, S., Nedeljkovic, M., Sumner, P., & Phillipou, A. (2022). Understanding self-report and neurocognitive assessments of cognitive flexibility in people with and without lifetime anorexia nervosa. *Cognitive neuropsychiatry*, *27*(5), 1–17. https://doi.org/10.1080/13546805.2022.2038554

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *21*(2), 169–183. https://doi.org/10.1055/s-2000-7563

Müller, U., & Kerns, K. (2015). The development of executive function. In L. S. Liben, U. Müller, & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science: Cognitive processes* (pp. 571–623). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy214

Nomi, J. S., Vij, S. G., Dajani, D. R., Steimke, R., Damaraju, E., Rachakonda, S., Calhoun, V. D., & Uddin, L. Q. (2017). Chronnectomic patterns and neural flexibility underlie executive function. *Neuroimage*, *147*, 861–871. https://doi.org/10.1016/j.neuroimage 2016.10.026

Oguz, G., Celikbas, Z., Batmaz, S., Cagli, S., & Sungur, M. Z. (2019). Comparison between obsessive compulsive disorder and panic disorder on metacognitive beliefs, emotional schemas, and cognitive flexibility. *International Journal of Cognitive Therapy*, *12*(3), 157–178. https://doi.org/10.1007/s41811-019-00047-5

Perpiñá, C., Segura, M., & Sánchez-Reales, S. (2017). Cognitive flexibility and decision-making in eating disorders and obesity. *Eating and Weight Disorders*, *22*(3), 435–444. https://doi.org/10.1007/s40519-016-0331-3

Rabin, L., Roth, R., Isquith, P., Wishart, H., Nutterupham, K., Pare, N., Flashman, L. A., & Saykin, A. (2006). Self- and informant reports of executive function on the BRIEF-A in MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology*, *21*(7), 721–732. https://doi.org/10.1016/j.acn.2006.08.004

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology*, *19*(5), 393–394. https://doi.org/10.1037/h0044509

Reitan, R. M. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual and motor skills*, *8*(3), 271–276. https://doi.org/10.2466/pms.1958.8.3.271

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. L., Lawrence, A. D., Mcinnes, L., & Rabbitt, P. M. A. (1998). A study of performance on test from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society*, *4*(5), 474–490. https://doi.org/10.1017/S1355617798455073

Roberts, M. E., Barthel, F. M. S., Lopez, C., Tchanturia, K., & Treasure, J. L. (2011). Development and validation of the Detail and Flexibility Questionnaire (DFlex) in eating disorders. *Eating Behaviors*, *12*(3), 168–174. https://doi.org/10.1016/j.eatbeh.2011.04.001

Rößner, A., Juniak, I., van Noort, B. M., Pfeiffer, E., Lehmkuhl, U., & Kappel, V. (2017). Cognitive flexibility in juvenile anorexia nervosa in relation to comorbid symptoms of depression, obsessive compulsive symptoms and duration of illness. *Zeitschrift Für Kinder- Und Jugendpsychiatrie Und Psychotherapie*, *45*(5), 371–390. https://doi.org/10.1024/1422-4917/a000493

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). Behavior rating inventory of executive function – adult version: Professional manual. *Psychological Assessment Resources*. https://doi.org/10.1109/icgse.2012 22

Settlage, P., Zable, M., & Harlow, H. F. (1948). Problem solutation by monkeys following bilateral removal of the prefrontal areas: VI. Performance on tests requiring contradictory reactions to similar and to identical stimuli. *Journal of Experimental Psychology*, *38*(1), 50–65. https://doi.org/10.1037/h0054430

Soto, E. F., Kofler, M. J., Singh, L. J., Wells, E. L., Irwin, L. N., Groves, N. B., & Miller, C. E. (2020). Executive functioning rating scales: Ecologically valid or construct invalid? *Neuropsychology*, *34*(6), 605–619. https://doi.org/10.1037/neu0000681

Spencer, R. J., Drag, L. L., Walker, S. J., & Bieliauskas, L. A. (2010). Self-reported cognitive symptoms following mild traumatic brain injury are poorly associated with neuropsychological performance in OIF/OEF veterans. *Journal of Rehabilitation Research and Development*, *47*(6), 521–530. https://doi.org/10.1682/JRRD.2009.11.0181

Stedal, K., & Dahlgren, C. L. (2015). Neuropsychological assessment in adolescents with anorexia nervosa – exploring the relationship between self-report and performance-based testing. *Journal of Eating Disorders*, *3*(27), 1–12. https://doi.org/10.1186/s40337-015-0062-2

Steinglass, J., Walsh, B. T., & Stern, Y. (2006). Set shifting deficit in anorexia nervosa. *Journal of the International Neuropsychological Society*, *12*(3), 431–435. https://doi.org/10.1017/S1355617706060528

Sternheim, L. C., van Passel, B., Dingemans, A., Cath, D., & Danner, U. N. (2022). Cognitive and experienced flexibility in patients with anorexia nervosa and obsessive compulsive disorder. *Frontiers in Psychiatry*, *13*, 868921. https://doi.org/10.3389/fpsyt.2022.868921

Strober, M. (1980). Personality and symptomatological features in young, nonchronic anorexia nervosa patients. *Journal of Psychosomatic Research*, *24*(6), 353–359. https://doi.org/10.1016/0022-3999(80)90027-6

Szmukler, G. I., Andrewes, D., Kingston, K., Chen, L., Stargatt, R., & Stanley, R. (1992). Neuropsychological impairment in anorexia nervosa: Before and after refeeding. *Journal of Clinical and Experimental Neuropsychology*, *14*(2), 347–352. https://doi.org/10.1080/01688639208402834

Tchanturia, K., Anderluh, M. B., Morris, R. G., Rabe-Hesketh, S., Collier, D. A., Sanchez, P., & Treasure, J. L. (2004). Cognitive flexibility in anorexia nervosa and bulimia nervosa. *Journal of the International Neuropsychological Society*, *10*(4), 513–520. https://doi.org/10.1017/S1355617704104086

Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., Treasure, J., & Morris, R. (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *Plos One*, *7*(1), e28331. https://doi.org/10.1371/journal.pone.0028331

Tchanturia, K., Lloyd, S., & Lang, K. (2013). Cognitive Remediation Therapy for anorexia nervosa: Current evidence and future research directions. *The International Journal of Eating Disorders*, *46*(5), 492–495. https://doi.org/10.1002/eat.22106

Tchanturia, K., Morris, R. G., Anderluh, M. B., Collier, D. A., Nikolaou, V., & Treasure, J. (2004). Set shifting in anorexia nervosa: An examination before and after weight gain, in full recovery and relationship to childhood and adult OCPD traits. *Journal of Psychiatric Research*, *38*(5), 545–552. https://doi.org/10.1016/j.jpsychires.2004.03.001

Tchanturia, K., Morris, R. G., Surguladze, S., & Treasure, J. (2002). An examination of perceptual and cognitive set shifting tasks in acute anorexia nervosa and following recovery. *Eating and Weight Disorders*, *7*(4), 312–315. https://doi.org/10.1007/BF03324978

Tchanturia, K., Serpell, L., Troop, N., & Treasure, J. (2001). Perceptual illusions in eating disorders: Rigid and fluctuating styles. *Journal of Behavior Therapy and Experimental Psychiatry*, *32*(3), 107–115. https://doi.org/10.1016/S0005-7916(01)00025-8

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. Teachers College, Columbia University.

Tokley, M., & Kemps, E. (2007). Preoccupation with detail contributes to poor abstraction in women with anorexia nervosa. *Journal of Clinical and Experimental Neuropsychology*, *29*(7), 734–741. https://doi.org/10.1080/13825580600966607

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Uddin, L. Q. (2021). Cognitive and beahavioural flexibility: Neural mechanisms and clinical considerations. *Nature Reviews*, *22*(3), 167–179. https://doi.org/10.1038/s41583-021-00428-w

Uznadze, D. (1966). *The psychology of set*. Consultants' Bureau.

Vall, E., & Wade, T. D. (2015). Trail Making Task performance in inpatients with anorexia nervosa and bulimia nervosa. *European Eating Disorders Review*, *23*(4), 304–311. https://doi.org/10.1002/erv.2364

Van Autreve, S., De Baene, W., Baeken, C., van Heeringen, K., Vancayseele, N., & Vervaet, M. (2016). Differential neural correlates of set-shifting in the bingeing– purging and restrictive subtypes of anorexia nervosa: An fMRI study. *European Eating Disorders Review*, *24*(4), 277–285. https://doi.org/10.1002/erv.2437

Verdejo-Garcia, A., Clark, L., Verdejo-Roman, Albein-Urios, N., Martinez-Gonzalez, M., Gutierrez, B., & Soriano-Mas, C. (2015). Neural substrates of cognitive flexibility in cocaine and gambling additions. *British Journal of Psychiatry*, *207*(2), 158–164. https://doi.org/10.1192/bjp.bp.114.152223

Wang, Y., Chen, J., & Yue, Z. (2017). Positive emotion facilitates cognitive flexibility: An fMRI study. *Frontiers in Psychology*, *8*, 1832. https://doi.org/10.3389/fpsyg.2017.01832

Wang, T., Li, M., Xu, S., Liu, B., Wu, T., Lu, F., Xie, J., Peng, L., & Wang, J. (2019). Relations between trait anxiety and depression: A mediated moderation model. *Journal of Affective Disorders*, *244*, 217–222. https://doi.org/10.1016/j.jad.2018.09.074

Weigl, E. (1927). On the psychology of so-called processes of abstraction. *The Journal of Abnormal and Social Psychology*, *36*(1), 3.

Wennerhold, L., & Friese, M. (2020). Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra: Psychology*, *6*(1), 9. https://doi.org/10.1525/collabra.276

Westwood, H., Stahl, D., Mandy, W., & Tchanturia, K. (2016). The set-shifting profiles of anorexia nervosa and autism spectrum disorder using the Wisconsin Card Sorting Test: A systematic review and meta-analysis. *Psychological Medicine*, *46*(9), 1809–1827. https://doi.org/10.1017/S0033291716000581

Wiemann, J. M. (1977). Explication and test of a model of communicative competence. *Human Communication Research*, *3*(3), 195–213. https://doi.org/10.1111/j.1468-2958.1977.tb00518 x

Yu, Y., Yu, Y., & Lin, Y. (2020). Anxiety and depression aggravate impulsiveness: The mediating and moderating role of cognitive flexibility. *Psychology, Health & Medicine*, *25*(1), 25–36. https://doi.org/10.1080/13548506.2019.1601748

Zhang, K., Fan, Z., Wang, Y., Faraone, S. V., Yang, L., & Chang, S. (2019). Genetic analysis for cognitive flexibility in the trail-making test in attention deficit hyperactivity disorder patients from single nucleutide polymorphism, gene to pathway level. *The World Journal of Biological Psychiatry*, *20*(6), 476–485. https://doi.org/10.1080/15622975 2017.1386324

CHAPTER SIX

Chapter 6 is currently under review:

Howlett, C.A., Stanford, T., Berryman, C., Karran, E.L., Bellan, V., Coussens, S., Miles, S., & Moseley, G.L. Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study. [Submitted 24th July 2023]

This article is not published, but under review; therefore, no permission is required to reproduce it in this thesis. A signed statement of contributions for this co-authored manuscript can be found in Appendix F.

Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study.

Caitlin A. Howlett, BPsych(Hons)[1], Tyman Stanford, PhD[2], Carolyn Berryman, PhD[1,3], Emma L. Karran, PhD[1], Valeria Bellan, PhD[1,4], Scott Coussens, PhD[4], Stephanie Miles, PhD[5,6,7], G. Lorimer Moseley, PhD[1]

1. Innovation, Implementation & Clinical Translation (IIMPACT) in Health, University of South Australia, Kaurna Country, Adelaide, Australia

2. Allied Health and Human Performance, University of South Australia, Kaurna Country, Adelaide, Australia

3. Brain Stimulation, Imaging and Cognition Research Group, School of Biomedicine, The University of Adelaide, Kaurna Country, Adelaide, Australia

4. Cognitive Neuroscience Laboratory – Australian Research Centre for Interactive and Virtual Environments, University of South Australia, Kaurna Country, Adelaide, Australia

5. Orygen, Melbourne, Australia

6. Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia

7. Department of Psychological Sciences, Swinburne University of Technology, Melbourne, Australia

Corresponding author: G. Lorimer Moseley, IIMPACT in Health, University of South Australia, GPO Box 2471, Kaurna Country, Adelaide 5001, South Australia, Australia.
Ph: +61 8 8302 1416
Email: lorimer.moseley@gmail.com

Abstract

Introduction: People with persistent pain experience problems modifying their cognition and behaviours when task or environmental demands change – abilities otherwise known as *cognitive flexibility*. However, limitations and inconsistent results of previous studies raise concerns over the quality of that evidence. We aimed to determine whether people with and without persistent pain differ on two assessments that are commonly used to assess cognitive flexibility. We also examined the relationship between the two assessments and explored whether people with and without persistent pain are distinguishable based on their scores on these assessments.

Methods: Participant demographics and symptoms of anxiety and depression were assessed. Participants completed the Cognitive Flexibility Inventory (CFI) and the Wisconsin Card Sorting Test (WCST). Multiple linear regression on the two outcome variables: CFI (total score) and WCST (% perseverative responses) was applied using backward stepwise selection. Both outcomes were calculated as a standardised proportion of the outcome scale and log-odds transformed to meet the model assumptions. Correlation analysis and logistic regression were used to investigate our secondary and exploratory aims.

Results: Data were available from 128 participants with persistent pain and 68 pain-free controls. After adjusting for covariates, no differences were found between people with and without persistent pain on either assessment of cognitive flexibility. No significant correlations were detected between the two assessments in either group. The probability of having persistent pain was also not associated with scores on either or both assessments.

Conclusion: 'Cognitive flexibility' appears similar in people with and without persistent pain.

Key words: Chronic pain; executive function; cognitive flexibility; neuropsychology; self-report

Introduction

Persistent pain, characterised by pain that exceeds the normal time frame required for bodily tissue to heal after injury (i.e., usually >3 months)[1,2], is a worldwide health problem.[3] Persistent pain affects approximately 20% of the population[4-6] and interferes with many aspects of everyday life including physical, emotional and social functioning.[7] Despite major advances in our knowledge and understanding of persistent pain, important gaps remain. Particularly, the role of cognitive factors in the development of and recovery from persistent pain, remains to be fully understood. Uncovering such factors has the potential to lead to more effective and tailored strategies to prevent and treat persistent pain.

One potentially important factor is what is commonly known as 'cognitive flexibility'. Cognitive flexibility is thought to be a multifaceted construct[8] that broadly refers to the ability to adjust thoughts, and cognitive and behavioural strategies when task or environmental demands change.[9,10] There are two approaches to assessing cognitive flexibility – self-report questionnaires and neuropsychological assessments. The Cognitive Flexibility Scale (CFS)[11], the Cognitive Flexibility Inventory (CFI)[12] and the Shift subscale of the Behavior Rating Inventory of Executive Function (BRIEF)[13] are commonly used self-report assessments of cognitive flexibility. Commonly used neuropsychological tests of cognitive flexibility include the Wisconsin Card Sorting Test (WCST)[14], the Trail Making Test (TMT)[15] and the Stroop Test.[16]

Several studies have investigated how people with persistent pain perform on neuropsychological assessments of cognitive flexibility, but those studies have yielded contrasting results. Some suggest people with persistent pain perform worse than pain-free controls;[17-19] others suggest they do not.[20-25] One study[26] found that people with fibromyalgia and people with musculoskeletal pain performed better on the Stroop Test (used here and by others e.g., Johnco et al.,[27], as a marker of cognitive flexibility) than pain-free controls, once depression and premorbid intelligence had been statistically controlled for. The mixed findings are difficult to interpret because most studies are underpowered[17,18,20,22,23,25,26], use a combination of assessments and outcome variables without correcting for multiple comparisons[18,22,24,26], or provide insufficient information about the task being used or how outcome variables were calculated.[19,23,26] Further, the assortment of neuropsychological assessments and outcome variables that are used to assess cognitive flexibility poses great challenges when comparing results across studies. When data are combined, meta-analyses reveal that people with persistent pain perform worse than pain-free controls on neuropsychological tests of cognitive flexibility[28-31], but high risk of bias of constituent

studies leaves some degree of uncertainty. Therefore, although the general pattern of systematic evidence indicates that people with persistent pain are impaired on neuropsychological assessments of cognitive flexibility, there are significant uncertainties around that evidence.

There is a building body of evidence that, although these self-report and neuropsychological assessments both purport to be capturing 'cognitive flexibility', they are unlikely to be tapping into shared, or similar, constructs.[32,33] In people with and without fibromyalgia, there appears to be a lack of association between self-reported cognitive and behavioural flexibility (i.e., Shift subscale of the BRIEF) and neuropsychological tests of cognitive flexibility (i.e., TMT B-A and WCST sum of perseverative responses).[21] Our two meta-analytic reviews also failed to detect a relationship between self-report questionnaires and neuropsychological tests of 'cognitive flexibility' in both clinical (11 studies)[33] and non-clinical (21 studies)[32] cohorts.

With the exception of two cross-sectional studies[21,34], no research to date has examined how people with persistent pain differ from pain-free controls when self-evaluating their ability to be cognitively and behaviourally flexible. Baker et al.[34] found that people with persistent pain reported worse functioning on the Shift subscale of the BRIEF than pain-free controls did, whereas Gelonch et al.[21] reported no differences between people with and without fibromyalgia on both self-report and neuropsychological assessments of cognitive flexibility, after adjusting for depression and anxiety.

It is possible that there may be distinguishable groups of people with persistent pain that can be identified on the basis of their performance on self-report and neuropsychological assessments of cognitive flexibility. Indeed, a study by Attal et al.[35] found that in a cohort of people undergoing surgical procedures for osteoarthritis and breast cancer, the presence of chronic post-surgical pain at 6 and 12-months was significantly predicted by performance on the TMT (part B). However, in people undergoing knee arthroplasty or noncardiac chest surgery, performance on the TMT (B-A) and Stroop (interference T-scores) did not predict the incidence of persistent post-surgical pain 6-months later.[36] To our knowledge, there have been no attempts to predict the presence of persistent pain based on concurrent self-report and neuropsychological assessments of 'cognitive flexibility'. It seems reasonable to suggest that there may be subgroups of people who are characterised by different patterns of self-reported and neuropsychological assessment results, but more evidence is needed to fill this gap in knowledge. Identifying such patterns may be important because if such groups exist, it would aid in the development of therapeutic approaches that are specifically targeted to individual needs.

The current study had three aims. Our first aim was to determine how people with persistent pain perform on a self-report questionnaire *and* on a neuropsychological assessment of cognitive flexibility relative to pain-free controls, after controlling for known confounders. We hypothesised that people with persistent pain would, on average, perform worse than their pain-free counterparts on both a self-report and a neuropsychological assessment of cognitive flexibility, even after controlling for confounders. Our second aim was to determine the relationship between self-reported cognitive flexibility and neuropsychologically assessed cognitive flexibility in people with and without persistent pain. Based on previous findings, we hypothesised that the two assessments would not relate. Our third, and exploratory aim, was to explore whether the probability of having persistent pain (or not) is associated with scores obtained on two concurrent cognitive flexibility assessments. No specific hypothesis was generated around this exploratory aim. For a full description of our *a priori* predictions, refer to the statistical analysis plan, which can be accessed via Open Science Framework (OSF) (https://osf.io/9q8ut/).

Materials and Methods

The University of South Australia's Human Research Ethics Committee (HREC) granted ethical approval for this online, cross-sectional study (approval no: 202439). The Checklist for Reporting Results of Internet E-Surveys (CHERRIES)[37] and the STROBE guidelines[38,39] were adhered to during the reporting of this study (see Supplemental Files 1 and 2). An *a priori* protocol was registered and locked under embargo on OSF (https://osf.io/9q8ut/). Deviations from the protocol or statistical analysis plan were recorded and were either outlined in this manuscript, or uploaded and time-stamped on OSF.

Participants

Participants from the general population were recruited using an online, convenience sampling strategy. The study was advertised via recruitment flyers, social media posts, announcements on the UniSA Research Volunteers website, and emails that were distributed using previously consented participant mailing-lists and the Research Participation Sign-up System, managed by the School of Psychology, Social Work & Policy, University of South Australia. Participants with or without persistent pain (i.e., pain experienced most days for three months or longer)[1,2] were invited to take part in an online survey. Participants were eligible if they were aged ≥18 years, had adequate ability to read and understand English, had normal or corrected-to-normal (e.g., glasses or contact lenses) vision, and had access to, and competency with, a computer/laptop or smartphone with internet access. Participants were excluded if they had colour vision deficiencies.

Measures

Primary outcomes

Self-reported cognitive flexibility – the Cognitive Flexibility Inventory (CFI)

The CFI is a 20-item scale that aims to assess: i) the tendency to realise that challenging situations are indeed controllable; ii) the ability to perceive different explanations for life incidences and human behaviour; and iii) the capability to produce multiple alternative solutions to difficult scenarios.[12] The CFI consists of two subscales: the Alternatives subscale (13 items) and the Control subscale (7 items). Respondents are required to indicate the degree to which they agree or disagree with various statements. Each item is rated on a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree), with some items being reversed scored. Higher scores on the CFI are considered to reflect greater self-reported cognitive flexibility. The CFI has previously demonstrated high internal consistency, adequate test-retest reliability over a 7-week time frame (r = .81, p <.001), and convergent validity with the Cognitive Flexibility Scale.[12] We selected the CFI because unlike other available questionnaires, it was developed with the intention of assessing components of cognitive flexibility that are thought to be adaptive when faced with stressful circumstances.[12] Here, the CFI's total score was used as an outcome of self-reported cognitive flexibility.

Neuropsychologically assessed cognitive flexibility – the Wisconsin Card Sorting Test (WCST)

The WCST[14] is a card-sorting task that has been widely used to assess cognitive flexibility.[27,40] Participants are instructed to 'match' 128 response cards, one by one, to one of four stimulus cards that vary in terms of their colour, form and number - see Fig 1. An electronic version of the WCST, which was scored in accordance with the standardised manual[14], was administered via Inquisit Web – an online software program hosted by Millisecond.[41] Participants were able to match a response card to one of the four multidimensional stimulus cards by clicking on it with either a computer mouse or with their fingertip via a touch-screen. Participants were not provided with direction on how to correctly sort the response cards, but relied only on written feedback ('correct' or 'wrong') displayed on the screen to determine the correct sorting principle with which to sort the cards by. After ten correct responses were made in succession, the correct sorting principle changed and participants were required to determine the new sorting principle through a process of trial and error. In line with the recommendations outlined by Miles et al.[42],

perseverative responses (percentage score) were used as an indicator of neuropsychological cognitive flexibility. Perseverative responses are continual responses made in accordance to a stimulus dimension that does not match the correct sorting principle i.e., the response is incorrect. However, since the cards are multidimensional and at times ambiguous – that is, a response card has the potential to match a stimulus card on multiple dimensions (e.g., colour and number) – perseverative responses may also be correct. A higher percentage of perseverative responses is thought to reflect poorer neuropsychologically assessed cognitive flexibility. We selected the WCST over other neuropsychological assessments because it most closely aligns with the definition of cognitive flexibility and is widely used in cognitive flexibility research. However, the validity and reliability of the WCST remains to be fully established. Sub-optimal psychometric properties of the WCST may partly stem from varying scoring procedures and administration modalities (i.e., manual or computerised), matters we have covered elsewhere.[42]

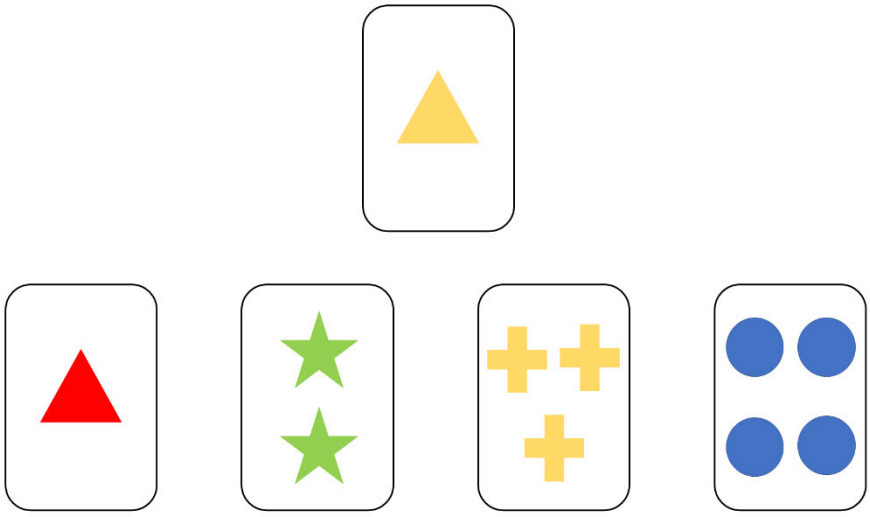


Figure 1. An illustrative example of the WCST. The single card with one yellow triangle is one of 128 response cards. The four bottom cards from the left-hand side to right hand side: one red triangle, two green stars, three yellow crosses and four blue circles are the stimulus cards.

Surveyed demographics

Participants responded to questions relating to age, sex assigned at birth, current gender identity, education level, current activity/occupation, financial situation, place of residence, country of residence, country of birth, refugee status, identification with ethnic or racial groups, mental

health, neurological conditions and medication use. Using adaptive questioning, all participants were then asked whether they have persistent pain, defined as *'pain experienced most days for 3 months or longer'*. Participants who responded *'yes'* to this question were presented with additional questions relating to pain intensity, duration, location(s), and diagnosis (if applicable).

Hospital Anxiety and Depression Scale (HADS)

The HADS is a questionnaire that is used to assess symptoms of anxiety and depression. The HADS consists of 14 items rated on a 4-point Likert scale ranging from 0-3. The 14 items are divided into two subscales: 7 items assess depressive symptoms and 7 items assess anxiety symptoms. The scores from each subscale range from 0-21, with higher scores reflecting more severe symptoms: 0-7 (normal), 8-10 (borderline abnormal), 11-21 (abnormal).[43,44] The two subscales of the HADS are internally consistent (HADS-D ranging from 0.67 to 0.90; HADS-A ranging from 0.68 to 0.93) and the HADS is a valid and reliable scale, with high sensitivity and specificity (range: 0.70 to 0.90).[45]

Procedure

The online survey was developed and administered using Inquisit Web.[41] Participants voluntarily accessed the online survey by following a link that prompted them to install the Inquisit 6 Player application (version 6.5.1) on their chosen device (either a computer or smartphone). Once Inquisit was installed, participants were instructed to click the *'start'* button, which directed them to an information sheet and consent form. Informed consent was obtained electronically from each participant by selecting the *'I wish to proceed'* button at the bottom of the page before commencing the study. The introductory page informed respondents about the purpose of the study, the eligibility criteria, the estimated time required for survey completion, the management of data and confidentiality, and that participation was not mandatory even after agreeing to take part. Participants were asked to enter a 6-digit alphanumerical ID code and were then asked to respond to 3 screening questions to confirm their eligibility: (1) are you at least 18 years of age? (2) are you colour-blind? and (3) can you read and understand English? Where participants were not eligible for this study, they were redirected to a page that notified them that they were ineligible to participate in this study and thanked them for their interest.

Eligible participants were asked whether they wished to receive a summary of the results and whether they would like to be contacted about future research being conducted – responding *'yes'* to either or both items prompted participants to enter their preferred email address in a text-box. Personal information (i.e., email addresses) were removed prior to data analysis and stored

separately from the data in a password protected file. Participants were then directed to a demographic questionnaire, self-report questionnaires and the WCST. Data for the Immune Status Questionnaire (ISQ)[46] were collected as part of a larger study, but were not analysed or reported here. The number of questionnaire items was limited to three per page. Questionnaire items were distributed across a different number of screen pages depending on group allocation – 41 pages for people with persistent pain and 36 pages for pain-free controls. Due to the nature of the WCST, presented screen pages varied across participants; the maximum number was 130 (including instruction and completion screen pages). It was compulsory to respond to every question that was presented in the survey, except for demographic questions relating to participants' life (e.g., age, education level, financial situation etc.); however, participants could quit the survey at any time by pressing *'Alt+E'* on their computer/laptop keyboard or by closing the application on their smartphone. The self-report questionnaires (i.e., CFI, HADS and ISQ) and the neuropsychological test (i.e., WCST) were presented in a counterbalanced order, such that some participants received these tests in the order presented above, while others received the tests in the reversed order (i.e., WCST, ISQ, HADS, CFI). The demographic questionnaire was always presented first. Participants were able to change their responses using a *'Back'* button up until the point of submitting each of the self-report questionnaires; it was not possible for participants to change their responses during the completion of the WCST.

Upon study completion (or where consent was withdrawn part way through the study), participants were directed to a Debriefing Form that outlined the true purpose and aims of the study. They were then provided with information relating to mental health services, instructions on how to uninstall Inquisit from their device and contact information of the project's primary investigator. Participants were given up to one week after study completion to withdraw their data from being included in the final analysis if they wished to do so. No honorariums were offered for study completion. A small group of colleagues within our laboratory piloted the survey, wherein they provided feedback on the clarity of the demographic items and the functionality of the survey as a whole. All recruitment materials and the full online survey are provided in the Supplementary material (see Supplemental Files 3-9).

Data analysis

Sample size calculation

Sample size calculations were made under the assumptions of the analysis to be performed on the two primary outcomes (see primary analysis for further details) to ensure sufficiently powered

primary analyses. Mean and standard deviation estimates of the outcome variables (WCST perseverative responses and CFI total score) in previous studies[12,27,47] were used. A detectable increase of 5% for WCST perseverative responses (%) in the persistent pain group than that of the pain-free control group was used, and a detectable decrease of 6 (similarly 5% of the scale) for the CFI (total score) in the persistent pain group than that of the pain-free control group was used. A required sample size of at least 120 would achieve 95% statistical power at the level of significance of 0.05. The sample size calculations were performed using simulation[48-50] of a beta distributed response of a linear model data generating process because both outcomes are approximately continuous, skewed and bounded. Sensitivity analysis of the required sample size was also performed using alternative model specifications and a range in the number of covariates and their relative effects. The empirical calculations were verified using analytical multiple linear regression power analysis methods available in G*Power 3[51,52] and the *pwr* package[53] in R (version 4.1.3).[54] Sample size calculations were undertaken prior to data analysis, but not prior to data collection. Full details regarding the sample size calculation can be found on OSF (https://osf.io/9q8ut/).

Statistical analysis plan

Data analysis was performed using R (version 4.2.2).[54] Raw data were imported into R using the *readr*[55] package. Data manipulation was performed using the *tidyverse*[56] and *janitor*[57] packages. Data visualisation and figures were created using *ggplot2*[58], *ggpubr*[59], *ggstatsplot*[60] and *jtools*[61], while diagnostic plots were generated using the *performance*[62] package. The *Classic Color Blind* colour palette via the *ggthemes*[63] package; the *Darjeeling1* colour palette via the *wesanderson*[64] package; the *Okabe-Ito* colour palette[65] via the *colorblindr*[66] package; and the base R *grDevices*[54] package were used in our attempt to make data visualisation more accessible to people who are colour-blind. Descriptive tables were created using *gtsummary*.[67] Linear and logistic regression outputs were interpreted using the *report*[68] package. Model equations were adapted from the *equatiomatic*[69] package. Other packages that were used include *writexl*[70], *skimr*[71], *car*[72], and *parameters*[73].

We decided *a priori* that i) the same user ID indicated duplicate participants and that we would retain the first entry for analysis and exclude any subsequent entries; ii) data collected prior to participant withdrawal (i.e., non-completion of primary outcomes) would not be analysed; and iii) data would be excluded from statistical analyses if ten or more identical responses were given in succession on the CFI. No data were missing for any of the variables used to perform our primary

analyses, so multiple imputation was not required (refer to the statistical analysis plan on OSF: https://osf.io/9q8ut/). Data for each of the reported non-binary identities did not meet our predetermined sample size (i.e., >5 participants). To protect the privacy and safety of participants, we needed to either i) combine non-binary gender identities into a third group (providing the reported non-binary gender identities were similar enough to warrant being combined) or ii) remove them from data analysis. Using The Trans Language Primer[74], we found that the reported non-binary identities were similar, so we decided to collapse them into a third group, namely *'transgender and non-binary individuals'*. Unless otherwise stated, statistical significance was set at $\alpha = 0.05$. Annotated R scripts that were used to undertake all aspects of data analysis can be accessed via OSF (https://osf.io/9q8ut/).

Primary analysis

Multiple linear regression

Separate regression analyses were carried out to i) determine whether people with persistent pain perform worse, on average, than pain-free controls on a neuropsychological assessment of cognitive flexibility and ii) determine whether self-reported cognitive flexibility is lower, on average, in those with persistent pain than in pain-free controls, after adjusting for confounding variables in both models. The following variables were entered into each model as covariates: age, gender (male, female, non-binary or transgender individuals), education level (high school, tertiary education, university higher degree), medication use (no, yes), HADS depression and anxiety scores, as well as test administration order ([order 1]: CFI first, WCST last vs. [order 2]: WCST first, CFI last), and modality of survey completion (computer/laptop vs. smartphone). The first level listed in each of the categorical covariates refers to the reference level used. Multiple linear regression with backwards stepwise selection was used to ensure model parsimony to avoid overfitting.[75] The predictor variable of interest was the grouping variable – that is, healthy control (HC) or persistent pain (PP); the outcome variables were the CFI (total score) and the WCST (% perseverative responses), calculated as proportions along their associated scales, herein, referred to as Cognitive Flexibility Inventory proportion of total score (CFI proportion) and WCST proportion of % perseverative responses (WCST proportion), respectively. These proportional outcome variables were transformed via log-odds transformation to satisfy the assumptions of multiple linear regression. Proportions equal to 0 or 1, which would result in undefined logit transformed values, were handled by adding (or subtracting) a half of the smallest unit of the scale. The assumptions of linearity, homoscedasticity of the residuals, influential observations,

normality of the residuals and multicollinearity were inspected. Initially, all predictor variables were entered into each regression model to create the full initial models. Predictors with *p*-values >0.05 were then removed in a stepwise manner – removing the predictor with the largest *p*-value first – until only significant predictors remained in both models.[76] Since the effect of group (i.e., PP or HC) on the two outcome variables (CFI and WCST proportion) was of primary interest, this variable remained in the final models irrespective of the *p*-value.

The aforementioned covariates were selected based on prior literature. Indeed, performance on the WCST has been found to be influenced by gender[77], and age and education.[14,78] Some studies have reported that poor cognitive performance is associated with depression and anxiety.[79,80] Further, medications including antidepressants, opioids, and anticonvulsants have been found to have both enhancing and detrimental effects on cognitive function.[81-83] On the other hand, gender[84], and symptoms of depression and anxiety have been found to influence scores on the CFI.[12,27,85,86] For the purposes of consistency, and that not every demographic characteristic has been previously investigated for its effects on both the CFI or WCST, we decided to enter each of these variables as covariates in the regression models of the two outcomes variables.

The full initial regression models for the $i = 1, 2\ldots, 196$ participants take the form:

$$\begin{aligned}\text{logit}(Y_i) = \beta_0 + \beta_1(\text{Group}_i = \text{Persistent pain}) + \beta_2(\text{Age}_i) \\ + \beta_3(\text{Educationlevel}_i = \text{Tertiary}) \\ + \beta_4(\text{Educationlevel}_i = \text{University higher degree}) \\ + \beta_5(\text{Gender}_i = \text{Female}) \\ + \beta_6(\text{Gender}_i = \text{Non-binary and/or transgender individuals}) \\ + \beta_7(\text{Medication}_i = \text{yes}) + \beta_8(\text{HADS_anxiety}_i) \\ + \beta_9(\text{HADS_depression}_i) + \beta_{10}(\text{Test order}_i = \text{Order 2}) \\ + \beta_{11}(\text{Modality}_i = \text{Smartphone}) + \epsilon_i\end{aligned}$$

where:

- Y_i is the outcome variable (a proportion of the outcome scale): either CFI total score or WCST % perseverative responses) for the i^{th} participant,
- β_0 is an intercept (the predicted value of $logit(Y_i)$ when all predictors are 0),
- β_j is the coefficient of the j^{th} predictor for $j = 1,2,\ldots,11$ (model coefficients),
- the categorical predictor variables take the form $(X_i = l)$ for level l of predictor X and takes the value 1 when $X_i = l$ for participant i (and 0 otherwise), and
- $\epsilon_i \sim N(0, \sigma)$ (independently and identically distributed normal error) for participant *i*.

Sensitivity analyses were carried out with univariate outliers – CFI or WCST proportion values that fell below Q1 - 1.5 × IRQ or exceeded Q3 + 1.5 × IQR – removed, which was visually confirmed via box-plots.

Secondary and exploratory analyses

Correlations

Currently, the relationship between self-reported and neuropsychological assessments of cognitive flexibility remains uncertain in people with persistent pain. To investigate the relationship between a self-report and neuropsychological assessment of cognitive flexibility, raw scores on the CFI subscales (i.e., Alternatives and Control) were correlated with raw perseverative response (%) scores on the WCST in people with and without persistent pain. Separate correlation analyses were performed for the persistent pain group and pain-free control group. Holm-Sidak adjusted significance thresholds (based on the sorted *p*-values) were applied to account for multiple comparisons and reduce the risk of Type I errors.[87] Holm-Sidak adjusted 95% confidence intervals of the estimated coefficients were also calculated. The assumptions of linearity or homoscedascity were independently assessed by two authors (CAH, TS), and a third author (GLM) was conferred with when consensus could not be reached. Effect sizes for correlation coefficients were considered small (0.1-0.29), medium (0.3-0.49), and large (>0.5) as recommended by Cohen[88]. Sensitivity analyses were carried out with univariate outliers – CFI (Alternatives and Control subscales) or WCST (% perseverative responses) values that fell below Q1 - 1.5 × IRQ or exceeded Q3 + 1.5 × IQR – removed, which was visually confirmed via box-plots.

Logistic regression

Logistic regression was used to explore the probability of the presence of persistent pain based on scores from a self-report and neuropsychological assessment of cognitive flexibility simultaneously. Predictor variables were centred and scaled based on sampled distributions of the observed values. A backward model reduction process was used, with non-significant model terms being removed in a stepwise manner.

The full initial logistic regression model was:

$$\text{logit}(p_i) = \beta_0 + \beta_1(\text{CFIproportion}_i) + \beta_2(\text{WCSTproportion}_i) + \beta_3(\text{CFIproportion}_i \times \text{WCSTproportion}_i)$$

$$Y_i \sim \text{Bernoulli}(p_i)$$

where:

- $i = 1,2,\ldots,196$ is the participant number,
- p_i is the probability of person i having persistent pain, that is, $p_i = P(\text{group}_i = \text{"persistent pain"})$ where group_i is the group ("persistent pain" or "healthy control") of person i and $P(x)$ is the probability of event x,
- β_0 is an intercept (the predicted value of $logit(Y_i)$ when all predictors are 0),
- $CFIproportion_i$ is the *CFI proportion of the total score* for the i^{th} participant (centered and scaled),
- $WCSTproportion_i$ is the *WCST proportion of % perseverative responses* for the i^{th} participant (centered and scaled), and
- Y_i is the observed PP status of person i, that is, $Y_i = 1$ if person i is in the group PP and $Y_i = 0$ if person i is not in the group PP (i.e., HC)

Results

Sample characteristics

A total of 241 participants (excluding duplicates) accessed the online survey between December 8th 2021 and April 9th 2022. Data from 196 participants were available for analysis (see Fig 2). Participant characteristics are summarised for each group separately using descriptive statistics in Table 1. Continuous variables are presented as values of central tendency (median) and dispersion (interquartile range). The minimum and maximum range of values for continuous variables are also provided. Dichotomous and categorical variables, as well as missing values, are presented as counts and percentages. Mental health conditions were broadly categorised by the primary investigator (CAH) using the Diagnostic Statistical Manual (DSM)-5-TR.[89] *Complex post-traumatic stress disorder* and *misophonia* were not otherwise categorised because they are not formally

recognised in the current version of the DSM.[89] Two experienced physiotherapists (GLM and CB) categorised pain diagnoses. Additional characteristics of the sample are provided in the Supplementary material (see Supplemental File 10).

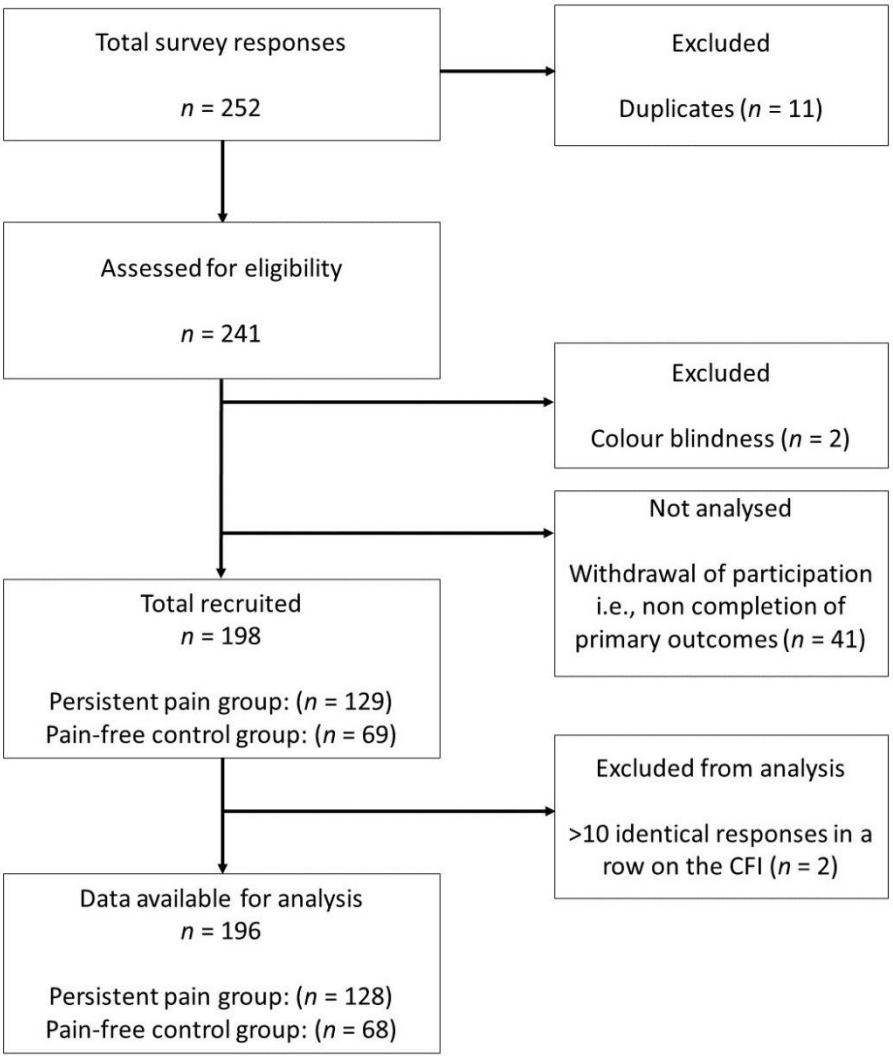


Figure 2. Participant flow diagram.

Table 1. Participant characteristics (N = 196).

Characteristic	**HC**, (n = 68)	**PP**, (n = 128)
Age (in years)		
Median (IQR)	36.0 (28.8, 44.2)	51.0 (38.0, 59.2)
Range	22 - 86	18 - 79
Sex		
Female	48 (71%)	105 (83%)
Male	20 (29%)	21 (17%)
Missing	-	2
Gender		
Female	48 (71%)	103 (80%)
Male	19 (28%)	20 (16%)
Non-binary and transgender individuals	1 (1.5%)	5 (3.9%)
Education level		
Secondary school (i.e., Year 12 or equivalent)	3 (4.4%)	16 (12%)
Tertiary education (e.g., Trade qualification/Certificate/Diploma/Bachelor's degree)	33 (49%)	72 (56%)
University higher degree (e.g., Masters, PhD)	32 (47%)	40 (31%)
Medications		
No	56 (82%)	63 (49%)
Yes	12 (18%)	65 (51%)
CFI (Alternatives subscale)		
Median (IQR)	74.0 (68.8, 78.0)	75.0 (69.0, 78.0)
Range	54 - 90	44 - 91
CFI (Control subscale)		
Median (IQR)	36.5 (31.0, 41.0)	35.0 (28.0, 40.0)
Range	14 - 49	8 - 49
CFI (total score)		
Median (IQR)	110.0 (100.8, 117.2)	109.5 (99.0, 117.0)
Range	73 - 139	54 - 135
WCST perseverative responses (sum)		
Median (IQR)	13.0 (7.0, 24.0)	18.5 (11.0, 27.0)
Range	3 - 76	3 - 120
WCST perseverative responses (%)		
Median (IQR)	11.9 (8.2, 19.6)	16.1 (11.1, 22.1)
Range	4 - 59	3 - 94

Characteristic	**HC**, (n = 68)	**PP**, (n = 128)
WCST perseverative errors (sum)		
Median (IQR)	12.0 (7.0, 23.0)	16.0 (10.0, 24.0)
Range	3 - 60	3 - 91
WCST perseverative errors (%)		
Median (IQR)	11.2 (8.2, 18.0)	14.3 (10.3, 19.5)
Range	4 - 47	3 - 71
HADS-Depression		
Median (IQR)	3.0 (2.0, 6.0)	7.0 (4.0, 10.2)
Range	0 - 13	0 - 19
HADS-Anxiety		
Median (IQR)	6.0 (3.0, 8.0)	9.0 (5.0, 12.0)
Range	0 - 17	0 - 20
Pain duration (in months)		
Median (IQR)	-	68 (24, 141)
Range	-	2 - 600
Missing	-	4
Pain intensity		
Median (IQR)	-	6 (4, 7)
Range	-	1 - 10
Missing	-	-
Pain location (primary)		
Abdominal	-	5 (4%)
Back	-	35 (27%)
Foot or leg	-	17 (13%)
Hand or arm	-	11 (9%)
Head	-	7 (5%)
Hip	-	10 (8%)
I find it difficult to identify my worst area of pain	-	24 (19%)
Neck	-	14 (11%)
Pelvic	-	5 (4%)
Pain location (secondary)		
Abdominal	-	24 (19%)
Back	-	60 (47%)
Foot or leg	-	57 (45%)
Hand or arm	-	58 (45%)
Head	-	35 (27%)
Hip	-	64 (50%)
Neck	-	65 (51%)

Characteristic	HC, (n = 68)	PP, (n = 128)
Pelvic	-	29 (23%)
Unknown	-	1 (1%)
Pain diagnosis		
Musculoskeletal	-	89 (70%)
Neuropathic	-	20 (16%)
Nociceptive	-	57 (45%)
Nociplastic	-	90 (70%)
Non-Musculoskeletal	-	12 (9%)
Unknown	-	1 (1%)

Note. CFI = Cognitive Flexibility Inventory; HADS = Hospital Anxiety and Depression Scale; HC = healthy control; IQR = interquartile range; PP = persistent pain; WCST = Wisconsin Card Sorting Test. Values and percentages for *pain location (secondary)* and *pain diagnosis* do not add up to the total *n* for the persistent pain group (128 or 100%) because participants could report having more than one secondary location of pain or their diagnosis could be classified into more than one category.

Group comparisons for assessments of cognitive flexibility

Multiple linear regression models were used to assess group differences on two outcomes of cognitive flexibility (CFI and WCST proportion), after accounting for the effects of the aforementioned covariates. Diagnostic plots were inspected for each of the backward selection models and no major deviations from statistical assumptions were observed (see Supplemental File 11).

Tables 2 and 3 show the final model reached from the final backward selection process for each outcome variable – CFI proportion and WCST proportion– respectively. No significant differences were found between people with and without persistent pain, after adjusting for covariates, on either assessment of cognitive flexibility: CFI proportion or WCST proportion. As shown in Table 2, the final model included the estimated effect of group, education level, and depression and anxiety on the log-odds values of the CFI proportion. The results indicated that 27.3% of the variance was explained by the model ($R^2 = 0.273$). Model comparisons revealed that the model goodness of fit (as assessed by the Akaike Information Criterion (AIC) value) improved during the backward selection process. No significant differences were found between groups on the log-odds scale of the CFI proportion ($\beta = 0.15$, 95% CI = -0.04, 0.35), after adjusting for education level and symptoms of depression and anxiety. This finding is consistent with the unadjusted median values (see Table 1). However, greater severity of anxiety symptoms was significantly associated with lower scores on the CFI proportion ($\beta = -0.04$, 95% CI = -0.06, -0.01), as was depressive symptoms ($\beta = -0.05$, 95% CI = -0.07, -0.02). Additionally, higher educational attainment was

associated with higher CFI proportion scores. Specifically, tertiary education (β = 0.33, 95% CI = 0.03, 0.63) and higher degree education (β = 0.52, 95% CI = 0.21, 0.83) were significantly different, on average, from high school education. A sensitivity analysis, with univariate outliers – defined as CFI proportion values that were more than 1.5 × IQR below Q1 or more than 1.5 × IQR above Q3 – removed from the final regression model did not change the statistical significance.

Table 2. Final regression model for the CFI (proportion of the total score).

Parameter	Coefficient	p-value	95% confidence interval
(Intercept)	1.22	0.000	(0.87, 1.57)
Group (persistent pain)	0.15	0.124	(-0.04, 0.35)
Education level (Tertiary)	0.33	0.031	(0.03, 0.63)
Education level (University higher degree)	0.52	0.001	(0.21, 0.83)
HADS anxiety score	-0.04	0.003	(-0.06, -0.01)
HADS depression score	-0.05	0.001	(-0.07, -0.02)

Note. CFI = Cognitive Flexibility Inventory; HADS = Hospital Anxiety and Depression Scale.

Table 3. Final regression model for the WCST (proportion of % perseverative responses).

Parameter	Coefficient	p-value	95% confidence interval
(Intercept)	-1.82	0.000	(-2.06, -1.58)
Group (persistent pain)	0.18	0.156	(-0.07, 0.43)
HADS anxiety score	-0.03	0.044	(-0.06, -0.00)
HADS depression score	0.04	0.011	(0.01, 0.08)

Note. HADS = Hospital Anxiety and Depression Scale; WCST = Wisconsin Card Sorting Test.

Table 3 shows the results of the same process applied to the WCST proportion log-odds outcome. The model was statistically significant (compared to an intercept model), but only explained 5.9% of the variance (R^2 = 0.059). Model comparisons revealed that the model goodness of fit (AIC value) improved throughout the backward selection procedure. No significant differences were observed between groups on the WCST proportion log-odds scale (β = 0.18, 95% CI = -0.07, 0.43). As shown in Table 1, the unadjusted median values appear to be different between the two groups, but this may be due to confounders. Greater severity of anxiety symptoms was significantly associated with better WCST proportion scores (β = -0.03, 95% CI = -0.06, -0.00), whereas greater severity of depressive symptoms was associated with worse WCST proportion scores (β = 0.04, 95% CI = 0.01,

0.08). A sensitivity analysis, with univariate outliers – defined as WCST proportion values that were more than 1.5 × IQR below Q1 or more than 1.5 × IQR above Q3 – removed from the final regression model resulted in scores of anxiety and depression no longer contributing significantly to the model. Graphical representations of the estimates for the final model parameters, as well as boxplots indicating outliers, can be found in the Supplementary material (see Supplemental Files 12-14).

Correlations between assessments of cognitive flexibility

The panel (CAH, TS, and GLM), which assessed the assumptions of linearity and homoscedascity deemed these assumptions to be violated by majority vote (see Supplemental File 15). Spearman's rho correlation coefficients were used. No significant correlations were found between either of the CFI subscales (i.e., Alternatives and Control) and the WCST (% perseverative responses) in the persistent pain group, nor in the pain-free control group (see Table 4). Sensitivity analyses, removing univariate outliers that were defined as WCST (% perseverative responses) and CFI (Alternatives and Control subscales) values that were more than 1.5 × IQR below Q1 or more than 1.5 × IQR above Q3, did not change the results in the pain-free control group, nor in the persistent pain group. Boxplots indicating outliers can be found in the Supplementary material (see Supplemental File 16).

Table 4. Spearman's rho correlation coefficients between self-reported cognitive flexibility and neuropsychological cognitive flexibility.

Group	Variables	rho	p-value (Holm-Sidak adjusted)	95% CI (Holm-Sidak adjusted)
Persistent pain	WCST (% perseverative responses) & CFI (alternatives)	-0.09	0.337	(-0.26, 0.09)
Persistent pain	WCST (% perseverative responses) & CFI (control)	0.01	1.000	(-0.21, 0.22)
Healthy control	WCST (% perseverative responses) & CFI (alternatives)	-0.10	0.818	(-0.36, 0.18)
Healthy control	WCST (% perseverative responses) & CFI (control)	-0.06	1.000	(-0.36, 0.27)

Note. CFI = Cognitive Flexibility Inventory; CI = confidence interval; rho = Spearman's rho; WCST = Wisconsin Card Sorting Test.

The probability of having persistent pain based on assessments of cognitive flexibility

A logistic regression model was fitted to explore whether the probability of having persistent pain is associated with two scores on assessments of cognitive flexibility (CFI proportion score and WCST proportion score). A backward selection process indicated that the most parsimonious model was the intercept only model (β = 0.63, 95% CI = 0.342, 0.932) (see Table 5). That is, the probability of having persistent pain was not associated with CFI proportion scores, nor WCST proportion scores. Refer to Supplemental File 17 for diagnostic plots for the logistic regression model.

Table 5. Model parameter estimates of each logistic regression model fit in the backward selection process.

Parameter	Model 1 (Interaction)	Model 2 (Main effects)	Model 3 (WCST model)	Model 4 (Intercept model)
(Intercept)	0.65 (0.36, 0.96)	0.65 (0.36, 0.96)	0.64 (0.35, 0.95)	0.63 (0.34, 0.93)
WCST	0.26 (-0.07, 0.65)	0.25 (-0.07, 0.62)	0.26 (-0.06, 0.63)	
CFI	-0.22 (-0.55, 0.1)	-0.2 (-0.51, 0.1)		
WCST x CFI	-0.08 (-0.53, 0.22)			

Statistics presented: coefficient estimate (95% confidence interval).
Note. CFI = Cognitive Flexibility Inventory; WCST = Wisconsin Card Sorting Test.

Discussion

Summary of findings

We investigated whether people with and without persistent pain differ in their ability to implement different cognitive and behavioural strategies when faced with changing environmental, or task demands. Contrary to our hypothesis, we found no evidence to suggest that people with persistent pain are different from pain-free controls on either a self-report questionnaire (i.e., CFI) or a neuropsychological test (i.e., WCST), after adjusting for covariates. However, we found that more severe symptoms of anxiety and depression were associated with lower scores on the CFI, and higher educational attainment was associated with higher CFI scores. This finding is consistent with prior research that has demonstrated that self-reported anxiety and depression are negatively related with self-reported cognitive flexibility.[12,27,85,86,90] Additionally,

once outliers were removed, we found no association between anxiety and depressive symptoms and performance on the WCST – a finding that is also consistent with previous research.[90]

Our study supports and extends the work of Gelonch et al.[21] who found that people diagnosed with fibromyalgia are not significantly different from pain-free controls on the Shift subscale of the BRIEF, after controlling for depression and anxiety. By contrast, Baker et al.[34] found that people with persistent pain reported impairments on the Shift subscale of the BRIEF relative to pain-free controls. However, unlike our study and that of Gelonch et al.[21], Baker et al.[34] did not control for the effects of depression and anxiety, and thus it is possible that the observed group differences in their sample could have been largely attributable to negative affect. Indeed, an inverse relationship between self-reported cognitive flexibility assessments and symptoms of depression and anxiety has been identified in people with and without comorbid depression and anxiety[27], in people with and without anorexia nervosa[90], and in undergraduate students[12], providing support for the potential influence of negative affect on self-reported cognitive flexibility. In line with our results, Gelonch et al.[21] also found that people with fibromyalgia were not significantly different from pain-free controls on the WCST (perseverative errors) after controlling for negative affect. Our findings also echo a number of previous studies that have found no differences between people with and without persistent pain on various neuropsychological assessments of cognitive flexibility.[20,22-25] However, our findings do not support those of large-scale meta-analyses that have demonstrated that people with persistent pain have small-to-moderate impairments on neuropsychological tests that are used to assess cognitive flexibility.[28-31,91]

So why then do previous meta-analyses show a small-to-moderate difference between people with and without persistent pain on neuropsychological tests that are commonly used to assess cognitive flexibility? A drawback of meta-analyses is that they, by default, rely on the included constituent studies to account for confounding factors (e.g., depression, medications etc.), which could potentially explain the discrepant findings between the current study and previous meta-analytical findings. It is possible that failing to adjust for important confounders in meta-analyses could give the false impression that people with persistent pain have an impaired ability to adapt and implement new strategies when task or environmental demands change, when in fact, these small-to-moderate effects could disappear once relevant confounding factors are controlled for. Further, despite power calculations for meta-analyses being recommended, they are rarely undertaken.[92] Indeed, it remains possible that previous meta-analyses could be underpowered, which could result in an overestimation of the true effect size, and in turn, misleading conclusions.[92,93]

We also aimed to investigate the strength and direction of the relationship between a self-report questionnaire and a neuropsychological assessment of cognitive flexibility in people with and without persistent pain. As expected, we did not find a significant relationship between our two chosen assessments – the CFI (Alternatives and Control subscales) and the WCST (% perseverative responses) in the persistent pain group, nor in the pain-free control group. Our findings corroborate previous research that indicates that self-report and neuropsychological assessments that purport to assess cognitive flexibility are poorly related. Specifically, Gelonch et al.[21] found no association between a self-report questionnaire (Shift subscale of the BRIEF) and multiple neuropsychological assessments of cognitive flexibility (TMT B-A and WCST perseverative responses) in people with and without fibromyalgia. Our results also support previous studies that have found little-to-no overlap between self-report and neuropsychological assessments of cognitive flexibility in non-clinical[27,90,94-96] and other clinical cohorts, including people with comorbid depression and anxiety[27], attention deficit hyperactivity disorder[95], anorexia nervosa[90,96], and obsessive-compulsive disorder.[97] Our group was the first to undertake two comprehensive meta-analyses that investigated the relationship between a range of self-report and neuropsychological assessments of cognitive flexibility. In both clinical[33] and non-clinical[32] cohorts, we found little-to-no relationship between self-report and neuropsychological assessments of cognitive flexibility. However, a recent study with a large sample (n = 246) of non-clinical, younger adults reported a small, but significant relationship (r = -0.22) between the CFS and the WCST (perseverative errors).[98] One potential explanation as to why our results conflict with that of Grant et al.[98] may be because they used different outcomes and had a much larger (and non-clinical) sample.

The lack of association between self-report and neuropsychological assessments found here and elsewhere is not confined to the construct of cognitive flexibility, but also extends to the executive function literature.[99] While it has been argued that strong correlations would be detectable if self-report and neuropsychological assessments assess the same construct[99], others have suggested that the non-correspondence between self-report and neuropsychological assessments should not be unexpected since the nature of the two assessments are fundamentally distinct – neuropsychological assessments often provide respondents with specific instructions, and thus are assumed to capture a person's maximum, momentary ability, whereas self-report questionnaires are thought to capture how respondents perceive their typical behaviours.[100] This may be true, but it seems unlikely to be the only difference between the assessment approaches likely to contribute. Our final and exploratory aim was to determine whether people with and

without persistent pain are distinguishable on the basis of their scores on both a self-report and a neuropsychological assessment. Our findings contradict that of Attal et al.[35], but are consistent with that of Vila et al.[36] because we found that the probability of having persistent pain (or not) was not associated with neuropsychological, nor self-reported cognitive flexibility.

Strengths and limitations

The strengths of this study include the preregistration of our protocol, as recommended in all pain research[101], our statistical analysis plan, our sufficiently powered sample, and that we controlled for confounders. Another strength of this study was that we limited the number of assessments and their respective outcomes, while also controlling for multiple comparisons where appropriate. Doing so addressed the shortcomings of previous studies and minimised the risk of Type I errors. Further, we adopted gender-inclusive research practices; we made the intentional decision to avoid categorising transgender and non-binary people into a single 'other' category because such categories reflect 'othering' of non-cisgendered people.[102,103] This study also has several limitations. First, IP addresses were not collected, so as to protect the anonymity of respondents. Therefore, it is possible that some participants may have completed the survey more than once using different participant ID codes, which may have resulted in duplicate data going unidentified. Although we cannot eliminate such a possibility, this scenario seems unlikely because participants did not receive incentives for their participation. Second, we did not exclude participants who were not naive to the WCST. Given that the WCST is susceptible to practice effects when re-administered within a 12-month time interval[104,105], prior knowledge of the rules that underpin the WCST could potentially yield better scores that are not reflective of enhanced cognitive flexibility, but rather measurement error.[106] Thus, there is a risk that such a decision could have skewed our results. Third, although we used a computerised version of the WCST that scores perseverative responses and perseverative errors in accordance to the Heaton et al.[14] manual, the scores obtained from computerised versions may not be psychometrically equivalent to the manual version.[107] Fourth, although the absence of a researcher during testing is likely to reduce elements of social desirability bias when completing the questionnaires, it meant that we were unable to monitor attention/distraction and other fidelity aspects when participants were undertaking the WCST. Fifth, the categorisation of medication use as a binary (yes/no) response has limitations – in particular, the response options did not allow us to capture dose or to examine whether certain medications have differential effects on the primary outcomes. We also did not assess or control for aspects of pain-related functioning (e.g., pain-related disability, quality of life or pain interference); future investigations using self-report or neuropsychological assessments to assess

cognitive flexibility in people with persistent pain conditions should consider controlling for such factors to better understand the association between persistent pain and the ability to adapt cognitive and behavioural strategies in response to changing task or environmental demands. Sixth, we relied on a convenience sampling strategy, so it is possible that our sample, and in turn our findings, were influenced by selection bias[108,109], and our sample was dominantly WEIRD (Western, Education, Industrialised, Rich and Democratic)[110], which limits the generalisability of our results. Lastly, we did not assess the validity or reliability of the items of the demographic questionnaire due to time constraints; however, the current study would have benefited from such an undertaking to ensure that the questions were precise enough to evoke consistent information from respondents[109].

Implications and considerations

Although there is sufficient evidence to support the notion that persistent pain is characterised by impairments in executive function more broadly[29,111,112], the findings of the current study suggest that the ability to adapt to new strategies in the face of changing demands, which is considered to be one aspect of executive function[8,113], may not be the component of executive function that is problematic in people with persistent pain. It remains possible that our chosen instruments were not sensitive enough to detect problems with self-reported and neuropsychologically assessed cognitive flexibility in people with persistent pain. However, that we found no difference between people with and without persistent pain on both self-report and neuropsychological assessments that purport to capture cognitive flexibility despite having a large sample size, controlling for multiple confounders, and restricting the number of assessments and outcome variables used, seems to challenge previous assertions that claim that more sensitive tests or advanced methods need to be developed to assess subtle distinctions in cognitive functioning.[114] Even if subtle impairments on the CFI and WCST do exist between people with and without persistent pain and more sensitive tests are required to detect such impairments, the effects are probably too small to be clinically meaningful.

It is also important to consider the drawbacks of both self-report and neuropsychological assessments of 'cognitive flexibility'. Self-report questionnaires that are designed to capture cognitive flexibility are likely influenced by the respondent's capacity to be self-aware and report the information in an accurate and non-biased manner. Similarly, a notorious problem of neuropsychological tests of 'cognitive flexibility' is that they inevitably rely on other executive and non-executive processes, which makes interpreting outcomes of such tasks difficult and rather

non-specific.[115,116] Although previous studies that have used the WCST report no differences in the number of perseverative errors between people with and without persistent pain conditions[24,25], the findings of other neuropsychological tests that are thought to assess cognitive flexibility – such as the TMT – are much harder to reconcile. Some studies have identified impaired performance on part B of the TMT in people with persistent low back pain[19] and fibromyalgia[18] relative to pain-free controls, while other studies have found no differences between people with and without persistent pain.[20,24,26] A potential explanation for these contrasting findings may be because the TMT - part B is fundamentally different from the WCST even though they are both used to assess the same construct of 'cognitive flexibility'. For instance, the TMT - part B requires respondents to alternate between an alpha-numeric sequence (i.e., 1-A-2-B-3-C etc.) as quickly as possible, whereas the WCST requires respondents to adapt to changing rules during card-sorting. Further, the outcome variables of the WCST and TMT are different: the TMT - part B is typically reported as time in seconds, whereas cognitive flexibility targeting outcomes of the WCST are most often presented as total sum or percentage scores (i.e., perseverative responses or perseverative errors). Although it has been suggested that part B of the TMT reflects, at least in part, cognitive flexibility[117], the influence of baseline motor speed, which is calculated by subtracting part A from part B[118], is seldom controlled for. Even when it is, difference scores do not produce adequate reliability coefficients[119,120]. For these reasons, future research may wish to steer anyway from relying solely on self-report and neuropsychological assessments to uncover problems in executive function and move towards using more ecologically valid tasks (e.g., the Virtual Cooking Task[121] or the Highway Driving Test[122]) that require individuals to perform tasks that are relevant to everyday life. Doing so could improve functional outcomes for people with persistent pain; however, it should be noted that such tasks need to be specifically developed and validated for people with persistent pain before their implementation.

It is important to note that our findings are not immediately applicable to what is termed *psychological flexibility*, which is the main construct targeted by Acceptance and Commitment Therapy (ACT).[123] Cognitive and psychological flexibility originated from different theoretical backgrounds; cognitive flexibility arose out of the neuropsychology field, whereas psychological flexibility has its roots in behavioural psychology, mindfulness and acceptance-based techniques. While there is some empirical evidence that self-reported cognitive flexibility (as assessed via the CFS) and self-reported psychological flexibility (as assessed via the Acceptance and Action Questionnaire-II (AAQ-II)) are related[98], the specific construct that underpins the AAQ-II has recently been put into question.[124,125] Thus, our findings cannot be applied to the construct of

psychological flexibility. It is also important to note that although ACT is often used as an intervention to enhance psychological flexibility and facilitate improvement of pain-related functioning in people with persistent pain[126], currently no data exist on whether people with and without persistent pain differ on outcomes of psychological flexibility.

Conclusion

We found no evidence that people with persistent pain and pain-free controls differ on self-report and neuropsychological assessments that purport to assess 'cognitive flexibility'. Our study did not find evidence to support the notion that people with persistent pain are less effective than pain-free controls at adapting their cognitive and behavioural strategies when environmental or task demands change. Despite evidence that executive function is compromised in people with persistent pain, our results suggest that 'cognitive flexibility' is not how that compromise occurs. However, other confounding factors need to be considered in future research and ecologically valid assessments of 'cognitive flexibility' specifically, and executive function, more broadly, will need to be developed before definitive conclusions can be drawn.

Conflicting interests

GLM has received support from California Institutes of Health; Reality Health; ConnectHealth UK; Workers' Compensation Boards in Australia, Europe and North America; and AIA Australia. Professional and scientific bodies have reimbursed him for travel costs related to presentation of research at scientific conferences/symposia. He has received speaker fees for lectures on pain and rehabilitation. He receives book royalties from NOIgroup publications, Dancing Giraffe Press & OPTP. CB has received support from Workers' Compensation Boards in Australia and Kaiser Permanente. All remaining authors (CAH, TS, ELK, VB, SC, and SM) declare no potential conflicts of interest (personal or financial) with respect to the research, authorship, or publication of this manuscript.

Funding

CB, ELK and GLM were supported by a National Health and Medical Research Council (NHMRC) Leadership Investigator Grant [ID 1178444] to GLM. CB was supported by a NHMRC Early Career Fellowship [ID 1127155]. ELK was supported by a project grant from Lifetime Support Authority of South Australia. These funding bodies had no role in the study design, collection, analysis or interpretation of the data, writing of the manuscript, or the decision to submit the manuscript for publication.

Ethical approval

Ethical approval was granted by the University of South Australia Human Research Ethics Committee (approval number: 202439).

Informed consent

Informed consent was obtained electronically from all subjects prior to study involvement.

Registration

The study protocol and statistical analysis plan were pre-registered on Open Science Framework (https://osf.io/9q8ut/).

Guarantor

CAH is the guarantor of this article.

Author contributions

CAH, CB, ELK, VB and GLM were involved in study conception. All authors were involved in protocol development. CAH researched literature, gained ethical approval, recruited subjects and performed data collection, and wrote the first draft of the manuscript. CAH, TS and GLM were involved in data analysis and/or interpretation. All authors reviewed, edited and approved the final version of the manuscript.

Data availability statement

All data that support the findings of this study are openly available on Open Science Framework (OSF) at https://osf.io/9q8ut/.

Orcid iDs

Caitlin A. Howlett: 0000-0002-4584-8641
Tyman Stanford: 0000-0002-8570-5493
Carolyn Berryman: 0000-0002-5316-0847
Emma L. Karran: 0000-0002-6088-8289
Valeria Bellan: 0000-0002-8512-5288
Scott Coussens: 0000-0002-3564-5859
Stephanie Miles: 0000-0001-8190-894X
G. Lorimer Moseley 0000-0002-3750-4945

References

1. Treede R-D, Rief W, Barke A, et al. A classification of chronic pain for ICD-11. *Pain* 2015; 156: 1003–7.

2. Merskey H and Bogduk N. *Classification of chronic pain: Descriptions of chronic pain syndromes and definitions of pain terms.* 2nd ed. Seattle, WA, USA: IASP Press, 1994.

3. James SL, Abate D, Abate KH, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; 392: 1789–858.

4. Pitcher MH, Von Korff M, Bushnell MC, et al. Prevalence and profile of high-impact chronic pain in the United States. *J Pain* 2019; 20: 146–60.

5. Sá KN, Moreira L, Baptista AF, et al. Prevalence of chronic pain in developing countries: Systematic review and meta-analysis. *Pain Reports* 2019; 4: e779.

6. Yong RJ, Mullins PM and Bhattacharyya N. Prevalence of chronic pain among adults in the United States. *Pain* 2022; 163: e328–32.

7. Hadi MA, McHugh GA and Closs SJ. Impact of chronic pain on patients' quality of life: A comparative mixed-methods study. *J Patient Exp* 2019; 6: 133–41.

8. Dajani DR and Uddin LQ. Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends Neurosci* 2015; 38: 571–8.

9. Cañas J, Quesada J, Antolí A, et al. Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics* 2003; 46: 482–501.

10. Scott WA. Cognitive complexity and cognitive flexibility. *Sociometry* 1962; 25: 405–14.

11. Martin MM and Rubin RB. A new measure of cognitive flexibility. *Psychol Rep* 1995; 76: 623–6.

12. Dennis JP and Vander Wal JS. The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognit Ther Res* 2010; 34: 241–53.

13. Roth RM, Isquith PK and Gioia GA. *Behavior Rating Inventory of Executive Function – Adult Version*. Lutz, Florida: Psychological Assessment Resources, Inc, 2005.

14. Heaton RK, Chelune GJ, Talley JL, et al. *Wisconsin Card Sorting Test manual: Revised and expanded.* Psychological Assessment Resources, 1993.

15. Reitan RM. The relation of the Trail Making Test to organic brain damage. *J Consult Psychol* 1955; 19: 393–4.

16. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol* 1935; 18: 643–62.

17. Lee D-H, Lee K-J, Cho KIK, et al. Brain alterations and neurocognitive dysfunction in patients with complex regional pain syndrome. *J Pain* 2015; 16: 580–6.

18. Tesio V, Torta DME, Colonna F, et al. Are fibromyalgia patients cognitively impaired? Objective and subjective neuropsychological evidence. *Arthritis Care Res* 2015; 67: 143–50.

19. Weiner DK, Rudy TE, Morrow L, et al. The relationship between pain, neuropsychological performance, and physical function in community-dwelling older adults with chronic low back pain. *Pain Med* 2006; 7: 60–70.

20. Cherry BJ, Zettel-Watson L, Shimizu R, et al. Cognitive performance in women aged 50 years and older with and without fibromyalgia. *Journals Gerontol Ser B Psychol Sci Soc Sci* 2014; 69: 199–208.

21. Gelonch O, Garolera M, Valls J, et al. Executive function in fibromyalgia: Comparing subjective and objective measures. *Compr Psychiatry* 2016; 66: 113–22.

22. Moriarty O, Ruane N, O'Gorman D, et al. Cognitive impairment in patients with chronic neuropathic or radicular pain: An interaction of pain and age. *Front Behav Neurosci* 2017; 11.

23. Nadar MS, Jasem Z and Manee FS. The cognitive functions in adults with chronic pain: A comparative study. *Pain Res Manag* 2016; 2016: 1–8.

24. Suhr JA. Neuropsychological impairment in fibromyalgia: Relation to depression, fatigue, and pain. *J Psychosom Res* 2003; 55: 321–9.

25. Verdejo-Garcia A, Lopez-Torrecillas F, Calandre EP, et al. Executive function and decision-making in women with fibromyalgia. *Arch Clin Neuropsychol* 2009; 24: 113–22.

26. Walitt B, Roebuck-Spencer T, Bleiberg J, et al. Automated neuropsychiatric measurements of information processing in fibromyalgia. *Rheumatol Int* 2008; 28: 561–6.

27. Johnco C, Wuthrich VM and Rapee RM. Reliability and validity of two self-report measures of cognitive flexibility. *Psychol Assess* 2014; 26: 1381–7.

28. Bell T, Trost Z, Buelow MT, et al. Meta-analysis of cognitive performance in fibromyalgia. *J Clin Exp Neuropsychol* 2018; 40: 698–714.

29. Berryman C, Stanton TR, Bowering KJ, et al. Do people with chronic pain have impaired executive function? A meta-analytical review. *Clin Psychol Rev* 2014; 34: 563–79.

30. Kessels RPC, Aleman A, Verhagen WIM, et al. Cognitive functioning after whiplash injury: A meta-analysis. *J Int Neuropsychol Soc* 2000; 6: 271–8.

31. Rathbone M, Parkinson W, Rehman Y, et al. Magnitude and variability of effect sizes for the associations between chronic pain and cognitive test performances: A meta-analysis. *Br J Pain* 2016; 10: 141–55.

32. Howlett CA, Wewege MA, Berryman C, et al. Same room - different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clin Psychol Rev* 2021; 88: 102061.

33. Howlett CA, Wewege MA, Berryman C, et al. Back to the drawing board – the relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts. A systematic review and meta-analysis. *Neuropsychology* 2022; 36: 347–72.

34. Baker KS, Gibson S, Georgiou-Karistianis N, et al. Everyday executive functioning in chronic pain: Specific deficits in working memory and emotion control, predicted by mood, medications, and pain interference. *Clin J Pain* 2016; 32: 673–80.

35. Attal N, Masselin-Dubois A, Martinez V, et al. Does cognitive functioning predict chronic pain? Results from a prospective surgical cohort. *Brain* 2014; 137: 904–17.

36. Vila MR, Todorovic MS, Tang C, et al. Cognitive flexibility and persistent post-surgical pain: The FLEXCAPP prospective observational study. *Br J Anaesth* 2020; 124: 614–22.

37. Eysenbach G. Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res* 2004; 6: e34.

38. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Int J Surg* 2014; 12: 1500–24.

39. von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* 2007; 370: 1453–7.

40. Tchanturia K, Davies H, Roberts M, et al. Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PLoS ONE* 2012; 7: e28331.

41. Millisecond Software. Inquisit 6, http://www.millisecond.com (2021).

42. Miles S, Howlett CA, Berryman C, et al. Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behav Res Methods* 2021; 53: 2083–91.

43. Stern AF. The Hospital Anxiety and Depression Scale. *Occup Med* 2014; 64: 393–4.

44. Zigmond AS and Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; 67: 361–70.

45. Bjelland I, Dahl AA, Haug TT, et al. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002; 52: 69–77.

46. Wilod Versprille LJF, van de Loo AJAE, Mackus M, et al. Development and validation of the Immune Status Questionnaire (ISQ). *Int J Environ Res Public Health* 2019; 16: 4743.

47. Greve KW, Stickle TR, Love JM, et al. Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study. *Arch Clin Neuropsychol* 2005; 20: 355–64.

48. Arnold BF, Hogan DR, Colford JM, et al. Simulation methods to estimate design power: An overview for applied research. *BMC Med Res Methodol* 2011; 11: 94.

49. Green P and MacLeod C. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol* 2016; 7: 493–8.

50. Kontopantelis E, Springate DA, Parisi R, et al. Simulation-based power calculations for mixed effects modeling: ipdpower in Stata. *J Stat Softw* 2016; 74: 1–25.

51. Faul F, Erdfelder E, Lang A-G, et al. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007; 39: 175–91.

52. Faul F, Erdfelder E, Buchner A, et al. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods* 2009; 41: 1149–60.

53. Champely S. pwr: Basic functions for power analysis. R package version 1.3-0, https://cran.r-project.org/package=pwr (2020).

54. R Core Team. R: A language and environment for statistical computing, https://cran.r-project.org/. (2022).

55. Wickham H, Hester J and Bryan J. readr: Read rectangular text data. R package version 2.1.2, https://cran.r-project.org/package=readr (2022).

56. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *J Open Source Softw* 2019; 4: 1686.

57. Firke S. janitor: Simple tools for examining and cleaning dirty data. R package version 2.2.0, https://cran.r-project.org/package=janitor (2023).

58. Wickham H. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York, https://ggplot2.tidyverse.org (2016).

59. Kassambara A. ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0, https://cran.r-project.org/package=ggpubr (2020).

60. Patil I. Visualizations with statistical details: The 'ggstatsplot' approach. *J Open Source Softw* 2021; 6: 3167.

61. Long JA. jtools: Analysis and presentation of social scientific data. R package version 2.2.0, https://cran.r-project.org/package=jtools (2022).

62. Lüdecke D, Ben-Shachar MS, Patil I, et al. performance: An R package for assessment, comparison and testing of statistical models. *J Open Source Softw* 2021; 6: 3139.

63. Arnold JB. ggthemes: Extra themes, scales and geoms for 'ggplot2'. R package version 4.2.4, https://cran.r-project.org/package=ggthemes (2021).

64. Ram K and Wickham H. wesanderson: A Wes Anderson palette generator. R package version 0.3.6, https://cran.r-project.org/package=wesanderson (2018).

65. Okabe M and Ito K. Color universal design (CUD): How to make figures and presentations that are friendly to colorblind people, https://jfly.uni-koeln.de/color/ (2008).

66. McWhite CD and Wilke CO. colorblindr: Simulate colorblindness in R figures, https://github.com/clauswilke/colorblindr (2022).

67. Sjoberg DD, Whiting K, Curry M, et al. Reproducible summary tables with the gtsummary package. *R J* 2021; 13: 570–80.

68. Makowski D, Ben-Shachar MS, Patil I, et al. Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption, https://easystats.github.io/report/ (2020).

69. Anderson D, Heiss A and Sumners J. equatiomatic: Transform models into 'LaTeX' equations. R package version 0.3.1, https://cran.r-project.org/package=equatiomatic (2022).

70. Ooms J. writexl: Export data frames to Excel 'xlsx' format. R package version 1.4.2, https://cran.r-project.org/package=writexl (2023).

71. Waring E, Quinn M, McNamara A, et al. skimr: Compact and flexible summaries of data. R package version 2.1.5, https://cran.r-project.org/package=skimr (2022).

72. Fox J and Weisberg S. *An R companion to applied regression.* 3rd ed. Sage, https://socialsciences.mcmaster.ca/jfox/Books/Companion/ (2019).

73. Lüdecke D, Ben-Shachar MS, Patil I, et al. Extracting, computing and exploring the parameters of statistical models using R. *J Open Source Softw* 2020; 5: 2445.

74. The Trans Language Primer, https://translanguageprimer.com (accessed 30 June 2022).

75. Hastie T, Tibshirani R and Friedman J. Model assessment and selection. In: *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York, NY, 2009, p. 219–59.

76. McCarthy RV, McCarthy MM and Ceccucci W. Predictive models using regression. In: *Applying predictive analytics*. Springer, Cham, 2022, p. 87–121.

77. Boone KB, Ghaffarian S, Lesser IM, et al. Wisconsin Card Sorting Test performance in healthy, older adults: Relationship to age, sex, education, and IQ. *J Clin Psychol* 1993; 49: 54–60.

78. Rhodes MG. Age-related differences in performance on the Wisconsin Card Sorting Test: A meta-analytic review. *Psychol Aging* 2004; 19: 482–94.

79. McDermott LM and Ebmeier KP. A meta-analysis of depression severity and cognitive function. *J Affect Disord* 2009; 119: 1–8.

80. Kurita GP, De Mattos Pimenta CA, Braga PE, et al. Cognitive function in patients with chronic pain treated with opioids: Characteristics and associated factors. *Acta Anaesthesiol Scand* 2012; 56: 1257–66.

81. Prado CE, Watt S and Crowe SF. A meta-analysis of the effects of antidepressants on cognitive functioning in depressed and non-depressed samples. *Neuropsychol Rev* 2018; 28: 32–72.

82. Khera T and Rangasamy V. Cognition and pain: A review. *Front Psychol* 2021; 12: 1–1.

83. Higgins DM, Martin AM, Baker DG, et al. The relationship between chronic pain and neurocognitive function: A systematic review. *Clin J Pain* 2018; 34: 262–75.

84. Wang C, Zhang Z, Wiley JA, et al. Gender differences in pleasure: The mediating roles of cognitive flexibility and emotional expressivity. *BMC Psychiatry* 2022; 22: 1–8.

85. Wang T, Li M, Xu S, et al. Relations between trait anxiety and depression: A mediated moderation model. *J Affect Disord* 2019; 244: 217–22.

86. Yu Y, Yu Y and Lin Y. Anxiety and depression aggravate impulsiveness: The mediating and moderating role of cognitive flexibility. *Psychol Health Med* 2020; 25: 25–36.

87. Holm S. A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979; 6: 65–70.

88. Cohen J. Quantitative methods in psychology: A power primer. *Psychol Bull* 1992; 112: 155–9.

89. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders.* 5th ed. 2022. doi: 10.1176/appi.books.9780890425787

90. Lounes N, Khan G and Tchanturia K. Assessment of cognitive flexibility in anorexia nervosa - self-report or experimental measure? A brief report. *J Int Neuropsychol Soc* 2011; 17: 925–8.

91. Bunk S, Preis L, Zuidema S, et al. Executive functions and pain. *Zeitschrift für Neuropsychol* 2019; 30: 169–96.

92. Griffin JW. Calculating statistical power for meta-analysis using metapower. *Quant Methods Psychol* 2021; 17: 24–39.

93. Thorlund K, Imberger G, Walsh M, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis - A simulation study. *PLoS ONE* 2011; 6: e25491.

94. Gonzalez CA, Figueroa IJ, Bellows BG, et al. A new behavioral measure of cognitive flexibility. In: *Harris, D. (eds) Engineering psychology and cognitive ergonomics. Understanding human cognition. EPCE 2013. Lecture notes in computer science, vol 8019.* Berlin, Heidelberg: Springer, 2013, p. 297–306.

95. Kercood S, Lineweaver TT, Frank CC, et al. Cognitive flexibility and its relationship to academic achievement and career choice of college students with and without attention deficit hyperactivity disorder. *J Postsecond Educ Disabil* 2017; 30: 329–44.

96. Miles S, Nedeljkovic M, Sumner P, et al. Understanding self-report and neurocognitive assessments of cognitive flexibility in people with and without lifetime anorexia nervosa. *Cogn Neuropsychiatry* 2022; 27: 325–41.

97. Sternheim LC, van Passel B, Dingemans A, et al. Cognitive and experienced flexibility in patients with anorexia nervosa and obsessive compulsive disorder. *Front Psychiatry* 2022; 13: 868921.

98. Grant A and Cassidy S. Exploring the relationship between psychological flexibility and self-report and task-based measures of cognitive flexibility. *J Context Behav Sci* 2022; 23: 144–50.

99. Toplak ME, West RF and Stanovich KE. Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *J Child Psychol Psychiatry* 2013; 54: 131–43.

100. Wennerhold L and Friese M. Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra Psychol* 2020; 6: 9.

101. Lee H, Lamb SE, Bagg MK, et al. Reproducible and replicable pain research: A critical review. *Pain* 2018; 159: 1683–9.

102. Camerson JJ and Stinson DA. Gender (mis)measurement: Guidelines for respecting gender diversity in psychological research. *Soc Personal Psychol Compass* 2019; 13: e12506.

103. Fraser G. Evaluating inclusive gender identity measures for use in quantitative psychological research. *Psychol Sex* 2018; 9: 343–57.

104. Basso MR, Bornstein RA and Lang JM. Practice effects on commonly used measures of executive function across twelve months. *Clin Neuropsychol* 1999; 13: 283–92.

105. Basso MR, Lowery N, Ghormley C, et al. Practice effects on the Wisconsin Card Sorting Test-64 card version across 12 months. *Clin Neuropsychol* 2001; 15: 471–8.

106. McCaffrey RJ and Westervelt HJ. Issues associated with repeated neuropsychological assessments. *Neuropsychol Rev* 1995; 5: 203–21.

107. Steinmetz JP, Brunner M, Loarer E, et al. Incomplete psychometric equivalence of scores obtained on the manual and the computer version of the Wisconsin Card Sorting Test? *Psychol Assess* 2010; 22: 199–202.

108. Alessi EJ and Martin JI. Conducting an Internet-based survey: Benefits, pitfalls, and lessons learned. *Soc Work Res* 2010; 34: 122–8.

109. Ball HL. Conducting online surveys. *J Hum Lact* 2019; 35: 413–7.

110. Gosling SD, Sandy CJ, John OP, et al. Wired but not WEIRD: The promise of the Internet in reaching more diverse samples. *Behav Brain Sci* 2010; 33: 94–5.

111. Berryman C, Stanton TR, Bowering KJ, et al. Evidence for working memory deficits in chronic pain: A systematic review and meta-analysis. *Pain* 2013; 154: 1181–96.

112. Moriarty O, McGuire BE and Finn DP. The effect of pain on cognitive function: A review of clinical and preclinical research. *Prog Neurobiol* 2011; 93: 385–404.

113. Diamond A. Executive functions. *Annu Rev Psychol* 2013; 64: 135–68.

114. McGuire BE. Chronic pain and cognitive function. *Pain* 2013; 154: 964–5.

115. Miyake A, Friedman NP, Emerson MJ, et al. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cogn Psychol* 2000; 41: 49–100.

116. Müller U and Kerns K. The development of executive function. In: *L. S. Liben, U. Müller, & R. M. Lerner (eds.), Handbook of child psychology and developmental science: Cognitive* processes. John Wiley & Sons, Inc., 2015, p. 571–623.

117. Kortte KB, Horner MD and Windham WK. The Trail Making Test, Part B: Cognitive flexibility or ability to maintain set? *Appl Neuropsychol* 2002; 9: 106–9.

118. Vall E and Wade TD. Trail Making Task performance in inpatients with anorexia nervosa and bulimia nervosa. *Eur Eat Disord Rev* 2015; 23: 304–11.

119. Kopp B. Neuropsychologist must keep their eyes on the reliability of difference measures. *J Int Neuropsychol Soc* 2011; 17: 562–3.

120. Zorowitz S and Niv Y. Improving the reliability of cognitive task measures: A narrative review. *Biol Psychiatry Cogn Neurosci Neuroimaging.*

121. Chicchi Giglioli IA, Gálvez BP, Granados AG, et al. The virtual cooking task: A preliminary comparison between neuropsychological and ecological virtual reality tests to assess executive functions alterations in patients affected by alcohol use disorder. *Cyberpsychology, Behav Soc Netw* 2021; 24: 673–82.

122. Veldhuijzen DS, van Wijck AJM, Willie F, et al. Effect of chronic nonmalignant pain on highway driving performance. *Pain* 2006; 122: 28–35.

123. Hayes SC, Luoma JB, Bond FW, et al. Acceptance and commitment therapy: Model, processes and outcomes. *Behav Res Ther* 2006; 44: 1–25.

124. Tyndall I, Waldeck D, Pancani L, et al. The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *J Context Behav Sci* 2019; 12: 278–84.

125. Wolgast M. What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behav Ther* 2014; 45: 831–9.

126. Wicksell RK, Kemani M, Jensen K, et al. Acceptance and commitment therapy for fibromyalgia: A randomized controlled trial. *Eur J Pain* 2013; 17: 599–611.

CHAPTER SEVEN

Discussion

7.1 Brief overview

The overarching aim of this thesis was to investigate cognitive flexibility as a potential contributor to the vulnerability of persistent pain. However, preliminary searches of the published literature uncovered important gaps that needed to be addressed before such an investigation could commence. Despite cognitive flexibility having been widely researched for the past 60 years (Scott, 1962; Stewin, 1983), a lack of consensus among experts, confusion surrounding assessments, and unsubstantiated assumptions have persisted within the field. Consequently, a substantial part of this thesis was concerned with understanding and clarifying important issues in the cognitive flexibility field.

With these points in mind, this thesis aimed to: i) improve the current use of the WCST, particularly in terms of the inconsistencies that exist surrounding the key terminology of the task and the scoring and interpretation of two variables that are commonly used to index cognitive (in)flexibility i.e., perseverative errors and perseverative responses, ii) evaluate the relationship between self-report and neuropsychological assessments of cognitive flexibility in healthy cohorts, iii) evaluate the relationship between self-report and neuropsychological assessments of cognitive flexibility in clinical cohorts, iv) explore the historical underpinnings of when, why and how self-report and neuropsychological assessments of cognitive flexibility have become conflated, and ix) determine whether people with persistent pain perform worse on currently available assessments of cognitive flexibility than pain-free controls do.

The current chapter offers a summary of the main findings obtained from each scientific enquiry and discusses the broader strengths and limitations, unforeseen challenges and key learnings, and the implications of the findings obtained from the set of research studies included within this thesis. Lastly, some recommendations and directions for future research endeavours are proposed.

7.2 Summary of the main findings

The Wisconsin Card Sorting Test (WCST) (Heaton et al., 1993) is a renowned neuropsychological task that is widely used by researchers and clinicians to, purportedly, assess cognitive flexibility. Despite the popularity of the WCST, its scoring procedures have historically remained a source of confusion for many of its users (Flashman et al., 1991; Greve, 1993). Such confusion has led to the misinterpretation and problematic use of two of its outcome variables, namely perseverative errors and perseverative responses, as well as inconsistencies surrounding the key terminology of

the task. In an attempt to resolve these problems, Chapter 2 provides a critical review covering the major changes of the WCST since its inception, how the two main scoring methods (i.e., Grant & Berg, 1948; Heaton et al., 1993) differ from one another, and outlines the key definitions and terminology. The main finding of this chapter was that using different procedures for scoring perseverative responses and perseverative errors on the WCST influences the scores obtained for these two variables, which in turn, changes results and impacts their generalisability. In light of this finding, this thesis was able to establish recommendations for using and reporting the WCST as an assessment of cognitive flexibility in research and clinical settings. In particular, it was recommended that researchers and clinicians use the commercial, standardised Heaton et al. (1993) version of the WCST, but where this is not possible, it was advised that studies be transparent about the implemented administration and scoring procedures. Further, it was advised that both perseverative responses and perseverative errors should be reported when using the WCST as an assessment of cognitive flexibility. The findings of this thesis have the potential to lead to wholesale changes of the WCST on various online software programs.

In Chapters 3 and 4, the implicit assumption that self-report and neuropsychological tests of cognitive flexibility assess the same overarching construct was explored in both healthy and clinical cohorts. Under that assumption, self-report and neuropsychological assessments of cognitive flexibility should be well correlated. In two systematic reviews, however, this thesis found no significant relationship between self-report and neuropsychological tests that purport to assess cognitive flexibility. Hence, this thesis concluded that the key constructs of these assessments need to be re-evaluated by the field.

Although little-to-no relationship between self-report and neuropsychological assessments of cognitive flexibility has been reported in previous research (Gelonch et al., 2016; Gonzalez, Figueroa, Bellows, Rhodes, & Youmans, 2013; Johnco et al., 2014; Lounes et al., 2011), an assumption that these two assessment approaches target the same underlying construct has nonetheless prevailed, and in turn, has led to their conflation both conceptually and practically. The two systematic meta-analytical reviews presented in Chapters 3 and 4 clearly demonstrate that self-report and neuropsychological assessments that purport to assess cognitive flexibility are not interrogating the same underlying construct. Chapter 5 built on this discovery by critically reviewing the historical literature around these assessments and their use – when, how, and why the two approaches have become conflated. Thorough literature searching (via backward and forward citation tracking) discovered that such conflation seems to have arisen from research on eating disorders, and is based on intuitive, but empirically unsupported, claims. For instance,

because self-report and neuropsychological assessments of 'cognitive flexibility' have the same overarching label, clinicians and researchers have mistakenly assumed that these assessments capture the same underlying construct. This thesis encourages researchers and clinicians to i) refrain from using the two assessment approaches interchangeably, ii) be mindful of the limitations of both assessment approaches, and iii) avoid describing these two assessment approaches under the umbrella term of 'cognitive flexibility'. Instead, researchers and clinicians should be more precise about what it is they are interested in: self-report evaluations of the ability to adapt thoughts and/or behaviours, or performance on a particular neuropsychological task.

Chapter 6 presents an online, cross-sectional study investigating whether people with and without persistent pain differ on both a self-report (Cognitive Flexibility Inventory) and neuropsychological (WCST) assessment of 'cognitive flexibility'. People reporting persistent pain did not rate themselves as less able to adapt their cognitive and behavioural strategies, nor did they make more perseverative responses on the WCST than pain-free controls did. This empirical evidence suggests that self-reported and neuropsychologically assessed 'cognitive flexibility' do not seem to be important in offering resilience against, or increasing vulnerability to, persistent pain.

As a whole, this thesis has shown that i) the assessment of cognitive flexibility is problematic on several grounds, ii) the available evidence strongly suggests that the two assessment approaches to cognitive flexibility probably assess different things that are not related, iii) the conflation of these assessments and their purported underlying construct has occurred over time and can be linked to key moments in research, its reporting and its underlying assumptions, and iv) when controlling for likely confounders, people with persistent pain are not systematically different from their pain-free counterparts when assessed using either approach.

7.3 Implications

This thesis provides insights into the unclear nature of the overarching construct of 'cognitive flexibility'. In particular, it attempts to dismantle various unsubstantiated assumptions that have long prevailed in the field. The findings of this thesis have the potential to encourage both researchers and clinicians to address the lack of consensus that currently surrounds the construct of cognitive flexibility by increasing the precision with which it is defined and assessed, and have implications for how researchers and clinicians should proceed with using and interpreting assessments that purport to assess cognitive flexibility.

7.3.1 *Reducing inconsistency and ambiguity through precise terminology, definitions and language*

There is no consensus or clearly constrained definition for the term 'cognitive flexibility' (Cherry et al., 2021; Howlett et al., 2021; Ionescu, 2012). In addition, other terms that have been used interchangeably with cognitive flexibility include, for example, *mental flexibility* (Diamond, 2013) and *psychological flexibility* (e.g., Asali et al., 2020; Palm & Follette, 2011; see also Section 1.3.2). Consequently, the inability to arrive at a unified conceptualisation has led to varying terms and definitions being used throughout the literature, with the current understanding of cognitive flexibility being described as "a puzzle with scattered pieces" (Ionescu, 2012). Inconsistent terminology and the lack of definitional consensus concerning the construct of cognitive flexibility may be largely attributable to the fact that it is considered to be a *latent construct*, meaning that it is 'hidden' and cannot be directly observed (El-Den, Schneider, Mirzaei, & Carter, 2020). Regardless, a failure to establish standardised terminology and a unified definition of cognitive flexibility has the potential for researchers and clinicians to believe that they are referring to the same construct, when in fact, that may not be the case (Block, 2000 as cited in Lawson & Robins, 2021), creating further potential for conceptual confusion, misunderstanding and miscommunication in both research and clinical settings. Such discrepancies also make it difficult to decipher which descriptions do or do not reflect cognitive flexibility, and whether *cognitive, mental* and *psychological* flexibility can be considered as sibling constructs, or whether they are in fact entirely distinct.

A lack of precision is not exclusively confined to the conceptualisation and definition of cognitive flexibility, but also extends to the assessments that are commonly utilised to capture it. There is little agreement regarding the underlying constructs of neuropsychological assessments, which often leads to inconsistent classifications of such tests across the field (Chaytor & Schmitter-Edgecombe, 2003). Chapters 3 and 4 revealed that there is great variability in the assessments and outcomes that are selected as indicators of 'cognitive (in)flexibility', and at times, multiple tests and outcomes are used, which in many instances prevents the direct comparison of findings across studies and inflates the risk of false positives, respectively. The findings of the two meta-analytic reviews (Chapters 3 and 4) suggested that the constructs behind self-report and neuropsychological assessments that are often assumed, and used, to capture 'cognitive flexibility' are not identical. Thus, the fact that cognitive flexibility is poorly defined, and that the self-report and neuropsychological tests that purport to assess it are not related should make researchers and clinicians wary of such assessments; it currently remains unclear whether both,

one or neither assessment targets cognitive flexibility. As outlined in Chapter 5, these two assessment approaches are often conflated with one another and used interchangeably, presumably because researchers and clinicians have been operating under the misguided assumption that two instruments can be considered as equivalent assessments of a construct because they bear the same name – coined over a century ago as the *jingle fallacy* (Thorndike, 1913).

Taken together, the findings of this thesis point to a desperate need for the field to re-evaluate the central constructs that underlie self-report and neuropsychological assessments that purport to assess 'cognitive flexibility' and to better understand their psychometric properties. This will require collaborative efforts from researchers and clinicians in the field, perhaps in the form of crowd-sourced replication studies (Nosek, Spies, & Motyl, 2012). Such efforts might prove to be challenging since replications are not typically prioritised by funders or existing publishing models (Munafò et al., 2017; Nosek et al., 2012). However, replication studies will significantly contribute to uncovering the truth about the construct validity of assessments that supposedly assess cognitive flexibility. In the meantime, using terms precisely so as to more accurately reflect the purpose and outcome of these assessments, would be a step in the right direction. For instance, rather than labelling the WCST as an assessment of 'cognitive flexibility', it might be more appropriate to refer to this assessment in regards to the outcome of interest: i.e., the ability to change rules in card-sorting without perseveration. Using self-report and neuropsychological assessments of 'cognitive flexibility' interchangeably or describing these assessments in such a way that conflates them, or implies that these two approaches assess the same underlying construct should be avoided. Further, where past or future research relies on a singular assessment of 'cognitive flexibility' – that is, a self-report questionnaire or neuropsychological test – researchers and clinicians should be cautious not to extrapolate the findings of one assessment to the other, nor should the findings be interpreted to be representative of the broader construct of 'cognitive flexibility'. Assuming equivalence between these two assessment approaches would lead to a range of unwanted consequences for research and clinical practice, including reinforcement of the jingle fallacy; an increased risk of producing findings that are invalid and unreliable; wasted resources and government funding; and potential misunderstandings of clinical conditions.

7.3.2 *There is a need for increased transparency surrounding commercial and non-commercial versions of the Wisconsin Card Sorting Test*

Aside from the standardised WCST manual (Heaton et al., 1993), it would seem prudent for software developers to be transparent when designing either commercial or non-commercial versions of the WCST by making the administration and scoring procedures openly available, so that users can judge whether the scoring procedures conform to that of Heaton et al. (1993), or implement another scoring method. Further, when researchers and clinicians use the WCST it is important that i) they report which version of the task was administered (manual or computerised), ii) they report which scoring procedure was used, iii) use both perseverative responses and perseverative errors when assessing cognitive flexibility, and iv) if the standardised manual is used, they define and describe the key terminology in such a way that is consistent with Heaton et al. (1993). Indeed, Chapter 2 found that implementing different procedures for scoring perseverative responses and perseverative errors on the WCST prevents pooling and comparison of data across studies.

7.3.3 *Confounders: Are key pieces of the puzzle being missed?*

The findings of this thesis emphasise the importance of controlling for particular confounders when researchers use self-report and neuropsychological assessments that purportedly assess cognitive flexibility. In Chapter 6, it was found that more severe symptoms of anxiety and depression were associated with lower scores on the CFI. In addition, higher educational attainment was associated with higher scores on the CFI. For the WCST, however, no association was found between task performance and symptoms of anxiety and depression once outliers were removed. Further, only a small portion of the variance was explained (5.9%) by the WCST model, which suggests that additional confounders need to be considered and statistically controlled for in future research. Hence, considering and controlling for important confounders in the design and data analysis of research studies will facilitate progress in better understanding the true relationship between exposures and outcomes.

7.4 Strengths, limitations, caveats, and speculations

A detailed discussion of the strengths and limitations specific to each thesis chapter, or any identified in the context of prior research and clinical practice, are outlined in the corresponding manuscripts. This section will discuss some of the broader strengths, limitations, caveats, and speculations that were pertinent across studies.

7.4.1 *Could cognitive flexibility assessments be related after all?*

Although the findings of this thesis and a building body of evidence suggests that self-report and neuropsychological tests of cognitive flexibility do not share construct validity (Gonzalez et al., 2013; Grant & Cassidy, 2022; Johnco et al., 2014; Kercood, Lineweaver, Frank, & Fromm, 2017; Lounes et al., 2011; Miles et al., 2022), it is important to remain open to the possibility that a relationship between self-report and neuropsychological assessments of 'cognitive flexibility' could indeed exist, but has not been clearly detected. A key limitation of the systematic reviews (Chapters 3 and 4) was that *a priori* power calculations were not undertaken for the individual random-effects meta-analyses that were performed, nor are statistical software currently available to conduct *a priori* power calculations for multi-level meta-analytical approaches (Griffin, 2021). Further, correspondence with constituent study authors – that is, requests for raw or aggregate data – occurred during the midst of the COVID-19 pandemic, which meant that many authors were unable to contribute to the results of the two systematic reviews due to 'work-from-home' policies being in place, and in turn, having restricted access to data that were stored on university servers and not accessible off-site. Consequently, it is possible that the meta-analyses in the two systematic reviews missed data and did not detect a relationship between self-report and neuropsychological assessments that purport to assess cognitive flexibility, either because the relationship was too small to be detected with the data available, or because the missing data were systematically different. The large number of studies that were included and the pattern of data that were collected strongly suggests against these possibilities, but they cannot be entirely ruled out.

The focus of the systematic reviews undertaken in this thesis was on adults. Thus, studies and data that included participants under the age of 18 were excluded. It remains possible that self-report and neuropsychological assessments of cognitive flexibility are related in children and adolescents, although limited data in adolescents appears to suggest that this may not be the case (Herbrich et al., 2019; Stedal & Dahlgren, 2015). A systematic investigation is nonetheless warranted to determine whether the findings of this thesis replicate in those under 18 years of age.

7.4.2 *Compliance with Open Science practices: Transparency, replicability, and reproducibility*

A major strength of this work is that the studies contained within this thesis met contemporary standards of transparency and rigour. *Open Science* is a movement that refers to "the concepts of openness, transparency, rigour, reproducibility, replicability, and accumulation of knowledge, all

of which are considered fundamental features of the scientific endeavour" (Crüwell et al., 2018, p. 3). Implementing open science practices (e.g., open data and transparent reporting) facilitates reproducible and replicable research by enabling other scientists to better understand the ins-and-outs of the study methodology, as well as expand and build upon original findings. Comprehensive resources for how to adopt open science practices across every stage of the research process have been developed in the context of psychological (Crüwell et al., 2019; Kathawalla, Silverstein, & Syed, 2021; Klein et al., 2018) and pain science (Lee et al., 2018) fields, and where possible, every effort was made to implement open science practices throughout the duration of this thesis.

The protocols for the two systematic reviews (Chapters 3 and 4) were pre-registered on the International Prospective Register of Systematic Reviews (PROSPERO) and Open Science Framework (OSF), and the protocol for the online, cross-sectional study (Chapter 6) was pre-registered on OSF *a priori*. All study materials, anonymous and de-identified data, and analysis code for the systematic reviews and the online, cross-sectional study were uploaded to a data repository (i.e., OSF) and made publicly available. Relevant manuscript reporting checklists were used, including the PRISMA statement (Moher et al., 2009; Shamseer et al., 2015) for transparent reporting of systematic reviews and meta-analyses (Chapters 3 and 4), the CHERRIES checklist (Eysenbach, 2004) for reporting online surveys (Chapter 6), and the STROBE statement (Vandenbroucke et al., 2014) for reporting observational studies (Chapter 6). Deviations from the original study protocols were recorded in either the corresponding manuscripts or outlined in revised (and time-stamped) versions of the protocols. Additionally, conflicts of interest and funding sources were disclosed in all five manuscripts. For the systematic reviews (Chapters 3 and 4), all articles were screened, extracted and assessed for risk of bias by two reviewers. Additionally, a 'co-pilot system' (Reimer et al., 2019) was adopted for all stages of the analysis; two reviewers ensured that anonymised raw data had been manually entered correctly, and two reviewers independently conducted the statistical analysis and contributed to, and inspected, the analysis code to safeguard against potential human errors. For the online, cross-sectional study (Chapter 6), a statistical analysis plan was devised and made available on OSF and once again, data analysis scripts were checked for accuracy by implementing the 'co-piloting system' (Reimer et al., 2019).

7.4.3 *Drawbacks of various study designs*

The nature of cross-sectional study designs enables data collection to be timely and inexpensive, but clear drawbacks exist. A key limitation of the cross-sectional study (Chapter 6) within this

thesis lies in the methodology used, namely the conduction of an online survey to recruit participants and undertake data collection. Sampling bias – a possible consequence of non-random sampling methods – is a common issue for online surveys (Ball, 2019). Although Internet-based approaches offer immense promise in reaching non-WEIRD (Western, Educated, Industrialised, Rich and Democratic) samples and in turn, mitigating the biases that are common in traditional recruitment methods (Gosling, Sandy, John, & Potter, 2010), a recognised disadvantage is that certain groups of people will not be represented, including those who lack access to the internet and technology (e.g., those with low socioeconomic status, illiterate individuals or those who live in developing countries) (Ball, 2019). Participants recruited as part of Chapter 6 formed a predominantly WEIRD sample, and consequently, such a non-representative sample can limit the generalisability of the findings to the target population (Alessi & Martin, 2010; Ball, 2019). In a similar way, the study design of the second critical review (Chapter 5), which provided a historical account of published work, presents as a potential limitation. A systematic search and synthesis of the literature was not feasible because an *a priori* question or hypothesis could not be generated and because it was suspected that relevant literature may exist primarily in pre-database articles, books and lectures. Taking on this question using, for example, a Delphi-technique was beyond the scope of, and resources available for, this thesis. However, such an approach may have yielded more detailed information in relation to the origination of the conflation between cognitive flexibility assessments.

7.4.4 *Limited options for critical appraisal tools*

The goal of a systematic review is to answer a specific question by identifying, critically appraising and synthesising original research – either quantitatively or qualitatively. An important aspect of the systematic review process is evaluating the methodological and reporting quality of the body of evidence (Sanderson, Tatt, & Higgins, 2007) because systematic reviews are used to develop recommendations for end-users, and identifying the shortcomings of previous work helps guide future investigations. As indicated in the systematic reviews (Chapters 3 and 4), a large portion of the included studies were cross-sectional. One point not discussed in the manuscripts of both systematic reviews was the challenge of finding a suitable critical appraisal tool that could be used to evaluate the risk of bias in cross-sectional studies. Despite various critical appraisal tools being available for experimental, observational studies, and diagnostic accuracy studies, three systematic reviews have reported that validated quality assessment tools for cross-sectional studies are scarce (Page, McKenzie, & Higgins, 2018; Sanderson et al., 2007; Zeng et al., 2015). In recent years, the Appraisal tool for Cross-Sectional Studies (AXIS) was developed for the purpose

of assessing the quality and risk of bias of cross-sectional studies (Downes, Brennan, Williams, & Dean, 2016), but the validity and reliability of the tool has not yet been examined. Similarly, the National Heart, Lung, and Blood Institute (NHLBI) developed the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies, but the tool has not been standardised (NHLBI, 2014). That the inclusion of studies was not contingent on the research question, nor subsequent statistical analyses being aligned with that of the systematic reviews meant that many of the criteria of existing critical appraisal tools were not applicable or did not consider such nuances. The decision to adapt the two aforementioned critical appraisal tools to create a customised tool has both advantages and disadvantages. While the customised risk of bias form was tailored to the types of biases deemed most relevant to the questions being investigated in the systematic reviews, it remains possible that important criteria for assessing risk of bias were missed or overlooked. Further, because the critical appraisal tool was customised, it had not been validated against other tools.

7.4.5 *Inherent issues of self-report and neuropsychological assessments*

One limitation that is relevant to the field, more broadly, concerns the current assessments that are employed to assess cognitive flexibility. One potential problem of neuropsychological assessments is that they tap into a multitude of executive and non-executive cognitive processes during task completion (Miyake, Emerson, & Friedman, 2000). A consequence of this inherent problem of neuropsychological tests – otherwise known as the task impurity problem – is that it is difficult to extract the variance or 'signal' of interest from the systematic variance attributable to non-executive processes, and measurement error (random noise) (Miyake & Friedman, 2012). That neuropsychological tests do not assess cognitive flexibility in isolation not only makes it difficult to determine the processes that may be driving task performance, but also makes interpreting scores considerably difficult and imprecise (Miyake, Emerson, et al., 2000; Müller & Kerns, 2015). The same general issue is also applicable to self-report questionnaires because the ability to report real-world thoughts and behaviours relies on both executive and non-executive processes, including meta-cognitive awareness, as well as contextual factors (i.e., motivation, distractions, stress etc).

To further complicate matters, different modalities of neuropsychological assessments are usually not psychometrically equivalent. In particular, scores obtained on two commercial versions of the WCST – the manual version and the computerised version – cannot be considered comparable (Steinmetz, Brunner, Loarer, & Houssemand, 2010); studies investigating the equivalence of paper-

and-pencil and computerised versions of the TMT have also found non-equivalence between these two modalities (Drapeau & Bastien-Toniazzo, 2007). Some neuropsychological assessments also tend to have low internal and/or test-retest reliability scores (Dang, King, & Inzlicht, 2020; Miyake, Friedman, et al., 2000). In particular, different versions of the TMT, including the original version (Reitan, 1955), the D-KEFS version (Delis et al., 2001) and the Comprehensive TMT (Reynolds, 2002) do not provide acceptable reliability coefficients (Kopp, Rösser, & Wessel, 2008), and repeated administrations of the WCST have been found to induce practice effects (Basso, Bornstein, & Lang, 1999; Basso, Lowery, Ghormley, & Bornstein, 2001). Therefore, it is important that researchers and clinicians remain aware of the integral limitations of using self-report and neuropsychological assessments, and in turn, the conclusions that can be drawn from them.

7.5 Unforeseen challenges and associated learnings during the PhD

7.5.1 *Benefits and barriers of statistical and experimental software*

Identifying and selecting an appropriate statistical software program to use for data analysis poses a significant challenge for researchers across various scientific disciplines. An obvious advantage of point-and-click meta-analytical software programs such as RevMan (The Cochrane Collaboration, 2020), Meta-Essentials (Suurmond, van Rhee, & Hak, 2017), Jamovi, (Şahin & Aybek, 2019) MedCalc (MedCalc Statistical Software, 2020), and Comprehensive Meta-Analysis (Borenstein, 2022) is their user-friendly graphical interfaces. However, these programs are often limited in their functionality. While programs such as MedCalc and Comprehensive Meta-Analysis offer free trials for potential users, after which the purchase of a proprietary software license is required, such programs do not foster replicable research, nor do they conform to open science practices because they are closed-sourced (Zheng, 2021). The systematic reviews (Chapters 3 and 4) required the identification of meta-analytical software that not only permitted >100 correlational data points to be pooled, but simultaneously accounted for the dependency of effects. Additionally, high quality images of forest plots were necessary for publication, and many of the point-and-click software programs do not offer this capability with such large datasets. RStudio is a free and open-source integrated development environment for R – a programming language for statistical computation and graphical representation (R Core Team, 2022) – and the *metafor* package (Viechtbauer, 2010), which is an extension of the R language, has such capabilities. Therefore, R (R Core Team, 2022) was deemed to be the most appropriate statistical software with which to analyse the data, but it required extensive work to gain sufficient proficiency in coding, as well as understanding the statistical approach itself. Prior research has

recognised that learning a programming language and how to code is a considerably arduous undertaking (Kathawalla et al., 2021; Zheng, 2021).

Choosing an appropriate software program to use for online experiments is another challenging aspect of conducting research, given the trade-offs that are associated with such programs. Many neuropsychological tests that are used to investigate cognitive flexibility, and executive processes more broadly, are protected by copyright and thus, the scoring procedures and interpretations remain concealed from the public and wider scientific community. For Chapter 6, a decision was made to use the Inquisit Web software, which is hosted by Millisecond (Millisecond Software, 2021), to create the online survey because i) the Inquisit version of the WCST conforms to the Heaton et al. (1993) scoring method and ii) unlike many other web-based programs, the scoring methods of the tests included in the Millisecond test library are open for public scrutiny prior to purchasing a software license. When using the WCST, the standardised scoring procedure (i.e., Heaton et al., 1993) is recommended (Miles et al., 2021), but the fact that the PAR computer version is expensive (>$1,000 USD) and does not support web-based data collection indeed limits its useability. Creation of the online survey within the Inquisit Web platform (Millisecond Software, 2021) also proved to be time intensive and challenging because the software relies on coded scripts to be written, and run. Undoubtedly, the decisions to use Inquisit Web (Millisecond Software, 2021) and R and R Studio (R Core Team, 2022) for collecting and analysing the data contained in this thesis, respectively, led to substantial delays, and various adjustments to the thesis timeline had to be made. However, these decisions also improved the quality of the research and built the skillset of the candidate, and can thus be considered worthwhile.

7.5.2 *Extracting data for inclusion in systematic reviews*

Novel and statistically significant findings have a greater chance of being published in scientific journals than null findings or replications, lending science susceptible to false positives and publication bias (Munafò et al., 2017). For Chapters 3 and 4, articles were included regardless of whether the research questions of individual studies aligned with that of the two systematic reviews. It has been argued that such an approach might reduce the risk that results are biased because the success of publication is not contingent on the novelty or statistical significance of the findings (Lange, 2020). However, this decision posed other challenges, not least that 96 authors had to be contacted to request raw or aggregate data because they were not included in the publication or were not reported in the form required, or because sufficient information to determine eligibility was not reported.

7.5.3 *Impact of the COVID-19 pandemic*

The COVID-19 pandemic had direct and indirect impacts. Direct impacts were time lost due to sickness for the core research team and collaborators. Indirect impacts were difficulty of contacting authors for the two systematic reviews (96 authors in total) and their difficulty in accessing data and uncertainty about when they would return to their workplace during periods of isolation. The impact of this uncertainty was substantial because many authors would pledge to respond when they return to work, expecting that to be days or weeks away, only to have their return further delayed. As such, delays were eventually very long, but if they had been predicted to be so long, the research team would have implemented contingencies to mitigate the impact of the delays.

7.6 Directions for future research

The following section suggests considerations for future research and proposes some directions for how subsequent investigations might begin to address the additional research gaps that have been identified from this thesis in order to advance the field.

7.6.1 *Consistent terminology and a uniform definition of cognitive flexibility is required*

Reaching standardised terminology and a precise definition of cognitive flexibility is perhaps the most immediate priority for future research to address. The ways in which cognitive flexibility has previously been defined are plentiful, and no two definitions are exactly the same (Cherry et al., 2021; Howlett et al., 2021; Ionescu, 2012). The conceptual chaos is evident when we consider how cognitive flexibility has been previously described, for example,

"… cognitive flexibility (also called set-shifting, mental flexibility, or mental set shifting and closely linked to creativity)" (Diamond, 2013, p.136).

The scope note for the PsycINFO database defines 'cognitive flexibility' as: "the capacity for mental adaptability and objective appraisal", which shares some conceptual overlap with the definitions that have been summarised in previous reviews (Cherry et al., 2021; Howlett et al., 2021; Ionescu, 2012), particularly in terms of the ability to be 'adaptable'. However, the definition offered by PsycINFO falls short when we consider that i) it is nowhere near as comprehensive as other definitions that have been offered, for example: "One component of executive function is cognitive flexibility, which refers to the ability to shift to different thoughts or actions depending on situational demands" (Geurts, Corbett, & Solomon, 2009), or "cognitive flexibility: changing perspectives or approaches to a problem, flexibly adjusting to new demands, rules, or priorities (as

in switching between tasks)" (Diamond, 2013) and ii) appraisals are inherently subjective in nature (i.e., are not free from personal influence) and cannot be considered objective, as the definition offered by PsycINFO suggests.

To move the field forward, one recommendation of this thesis would be for future research to conduct a large-scale Delphi study that involves international experts, with the aim of reaching consensus on a precise definition of cognitive flexibility that can be universally adopted. The Collaboratory on Research Definitions for Reserve and Resilience in Cognitive Aging and Dementia has developed a unified framework for conceptually and operationally defining the construct of 'cognitive reserve' (https://reserveandresilience.com/framework/), which indicates that similar missions are not implausible for the construct of cognitive flexibility. Once a foundational definition has been established, it would seem vitally important for scientific databases such as PsycINFO to update their scope notes to include the newly conceived definition of cognitive flexibility. This action would also assist with appropriately indexing, and in turn, optimising searches, leading to the retrieval of relevant research articles within databases, as well as prevent researchers from becoming burdened by the conceptual murkiness of this field – an unresolved challenge of this thesis.

7.6.2 *Next on the agenda: Develop relevant and ecologically valid assessments of cognitive flexibility*

Improving existing assessments or developing new self-report questionnaires and ecologically valid tasks that are designed to assess cognitive flexibility with specific populations in mind is another important avenue for future research. One potential option may be to undertake an item content analysis to determine whether the items/content of existing self-report questionnaires that purport to assess cognitive flexibility are actually endorsed as assessments of cognitive flexibility, or whether the items/content of such questionnaires would be better captured by other theoretical constructs. Further, many neuropsychological tests that are commonly used to assess cognitive flexibility in research and clinical practice were not originally developed, or intended to be used, for such purposes. Rather, these assessments have historically been 'retrofitted' such that they have been relabelled as assessments of cognitive flexibility. For instance, the TMT was initially designed to assess cognitive dysfunction in people with brain injuries (Reitan, 1958), while the Stroop Colour-Word Test was originally created as an assessment of inhibition (Stroop, 1935). Further, outcome variables of particular assessments have also been retrofitted. As indicated in Chapters 2 and 5, the WCST was devised as an assessment of 'flexibility in thinking' or 'shifting of

set', however, the original version did not include outcome variables of perseveration; instead, the average number of errors made in each of the categories was used to denote (in)flexibility (Berg, 1948). It wasn't until the second iteration of the test that outcome variables such as *perseverative responses* and *perseverative errors*, variables which are now widely used to capture cognitive (in)flexibility, were formally introduced (Grant & Berg, 1948).

The most reasonable interpretation of the findings from Chapter 6 is that people with persistent pain are no less able to adapt cognitive and behavioural strategies on the CFI or WCST. The discussion section of Chapter 6 focuses on the fact that no differences on self-report and neuropsychological assessments of cognitive flexibility were observed between people with and without persistent pain when confounders were adequately controlled for. However, another interesting possibility is that the opposite could be true – that people with persistent pain do systematically differ from pain-free controls in their ability to adapt cognitive and behavioural strategies to changing demands, but the CFI and WCST lack the capacity to detect nuanced differences between these two groups. Regardless, the findings suggest that these two assessments are unlikely to offer any clinical utility or be very informative for devising new treatment targets. Indeed, it can be argued that pain persisting when it serves no protective function appears to resemble some form of 'inflexibility' within the nervous system, but it is unlikely that available assessments interrogate the exact processes that are responsible for ongoing pain, despite the absence of injury. This raises the interesting possibility that people with persistent pain conditions could indeed experience problems with being cognitively and behaviourally flexible, and in turn, lack the ability to implement new strategies, but that such subtle differences can only be observed when assessments include an element of personal salience – that is, related to pain or the perceived threat of pain. This might involve devising new assessments or adapting existing ones to determine whether an enhanced level of actual or perceived bodily threat impacts the ability to be cognitive and behaviourally flexible. One might ask: theoretically, what might a new or modified assessment that specifically targets the ability to adapt cognitive and behavioural strategies to changing demands look like? Developing self-report questionnaires that specifically ask questions in relation to the ability to adapt when experiencing a flare-up might be an appropriate next step. Similarly, incorporating noxious stimuli in immersive tasks that have real-world applications, including the *virtual cooking task* (Chicchi Giglioli, Gálvez, Granados, & Raya, 2021), could be a viable option, but its validity in people with persistent pain will first need to be verified.

7.6.3 *Is the standardised manual of the Wisconsin Card Sorting Test in need of another revision?*

Perhaps the field may benefit from another revision to the WCST manual, but with the particular intention to simplify the scoring procedures. The last comprehensive revision of the standardised WCST manual was in 1993 (Heaton et al., 1993). As demonstrated in Chapter 2, the WCST is a complicated task to interpret and score. Prior to the 1993 WCST standardised manual, some had attempted to clarify how to score perseverative responses and perseverative errors by creating supplemental scoring material (Axelrod et al., 1992; Flashman et al., 1991). The complexity of the WCST primarily stems from the fact that the perseverated-to principle (the principle that determines whether a response is scored as perseverative or not) is dictated by the examinee's responses, which means that the perseverated-to principle can change multiple times throughout the completion of the task. Further, different scoring procedures are applied depending on whether the responses are correct or incorrect, ambiguous or unambiguous, or perseverative or non-perseverative. For example, ambiguous responses must be 'sandwiched' between two unambiguous perseverative errors and match the persevered-to-principle that is currently in effect to be scored as a 'perseverative response'. Ultimately, this means that for the most part, participant responses cannot be scored in isolation, but must be considered in the context of all other previous and subsequent responses, making it arguably one of the most challenging tasks to comprehend, interpret and score.

The current version of the WCST manual (Heaton et al., 1993) presents the procedures for scoring perseverative responses and perseverative errors descriptively, along with supporting illustrations. However, presenting the scoring procedures in this manner is time-consuming for novice users (Axelrod et al., 1992). Although high inter-rater reliability has been found among novice users of the WCST, it takes approximately 143 minutes to learn how to score perseverative responses and perseverative errors when relying on the WCST instructions alone, but that number is reduced by 41% when supplemental instructions are used (Axelrod et al., 1992). With such complex scoring procedures, the potential for human errors among users of the WCST is difficult to ignore. To avoid the incidence of errors, Heaton et al. (1993) may consider incorporating schematic 'decision trees' as part of the WCST manual to assist users in accurately scoring complex responses (e.g., how to score the task when the perseverated-to principle changes, or how to score ambiguous responses). Such changes to the WCST manual will not only improve accuracy, but might also allow users of the task – particularly novices – to learn and score perseverative responses and perseverative errors with more ease.

Presenting the procedures for scoring perseverative responses in the form of diagrammatic 'decision trees' has been previously attempted by Berry (1996) (Fig 1-3); however, potential improvements of the original diagrams might include the following modifications (as shown in Fig 4):

i) combining Berry's (1996) three diagrams (i.e., Diagram 1a, Diagram 1b and Diagram 2 – refer to Figs 1-3) into one (Fig 4) so that users can see how these three schematic representations are interconnected
ii) removing any redundant information to reduce ambiguity and complexity
iii) using terminology that is consistent with the standardised manual (e.g., replacing "*perseveration principle"* with "*perseverated-to principle"*)
iv) expanding the diagram to also include diagrammatic instructions for how to score perseverative errors
v) providing a more comprehensive notes section so that users of the WCST can seek additional information about particular steps within the decision tree, ideally without the need to re-locate these key points within the main text of the WCST manual

It should be noted that this diagrammatic scoring procedure should not replace the standardised WCST manual (1993), and users are strongly encouraged to familiarise themselves with the WCST manual in the first instance. Further, it is intended to mirror the procedures for scoring perseverative responses and perseverative errors set out by Heaton et al. (1993), and only focuses on these two variables. Therefore, the standardised manual should be referred to for information concerning the administration of the WCST, or how to score other key variables of the task.

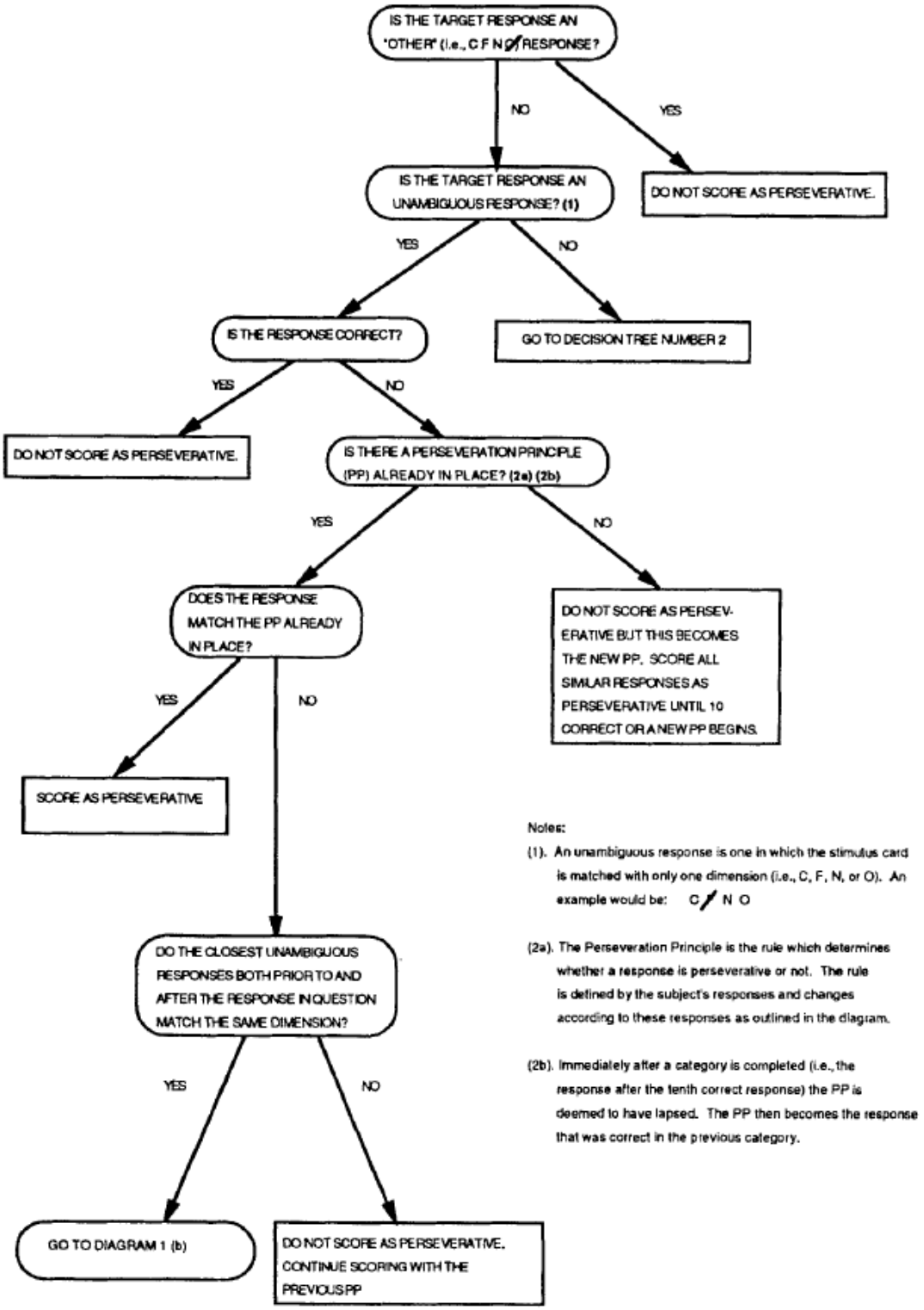
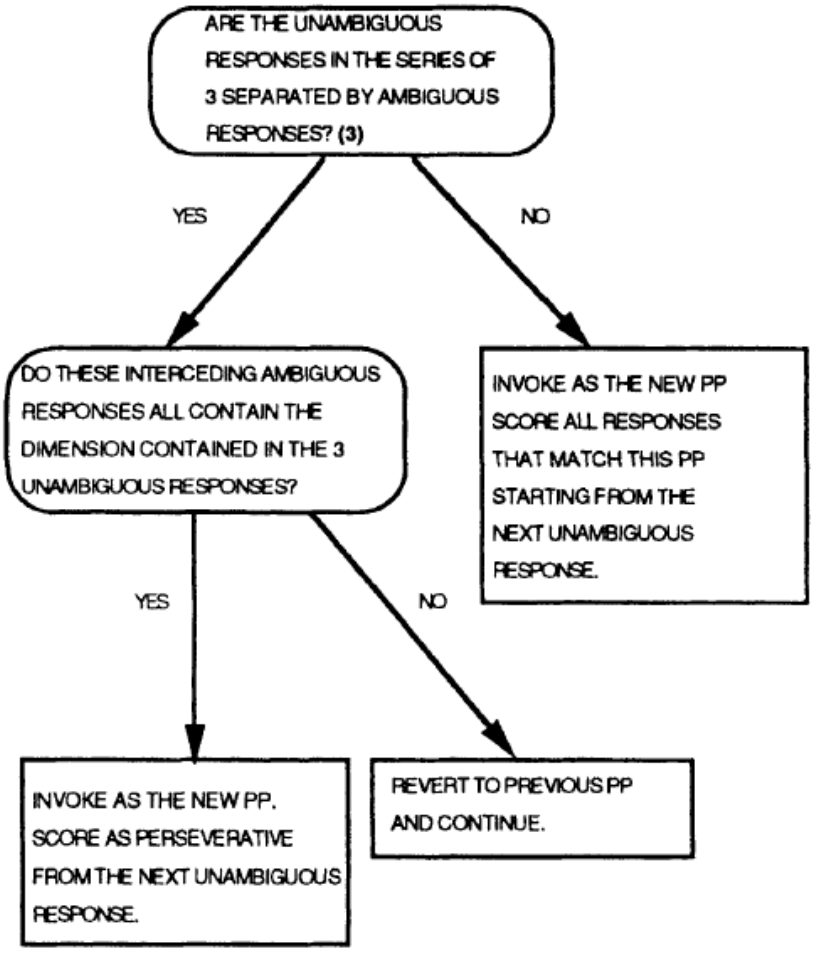


Figure 1. Steve Berry's original diagrammatic procedure for scoring unambiguous perseverative responses on the Wisconsin Card Sorting Test. Berry, S. Diagrammatic procedure for scoring the Wisconsin Card Sorting Test, *The Clinical Neuropsychologist*, © copyright [1996], reprinted with permission of Informa UK Limited, trading as Taylor & Taylor & Francis Group, http://www.tandfonline.com



Notes:

(3). An ambiguous response is one in which the stimulus card is matched with more than one dimension. An example is ✓✓✓O.

Figure 2. Steve Berry's original diagrammatic procedure for scoring unambiguous perseverative responses on the Wisconsin Card Sorting Test (continuation of Figure 1). Berry, S. Diagrammatic procedure for scoring the Wisconsin Card Sorting Test, *The Clinical Neuropsychologist*, © copyright [1996], reprinted with permission of Informa UK Limited, trading as Taylor & Taylor & Francis Group, http://www.tandfonline.com

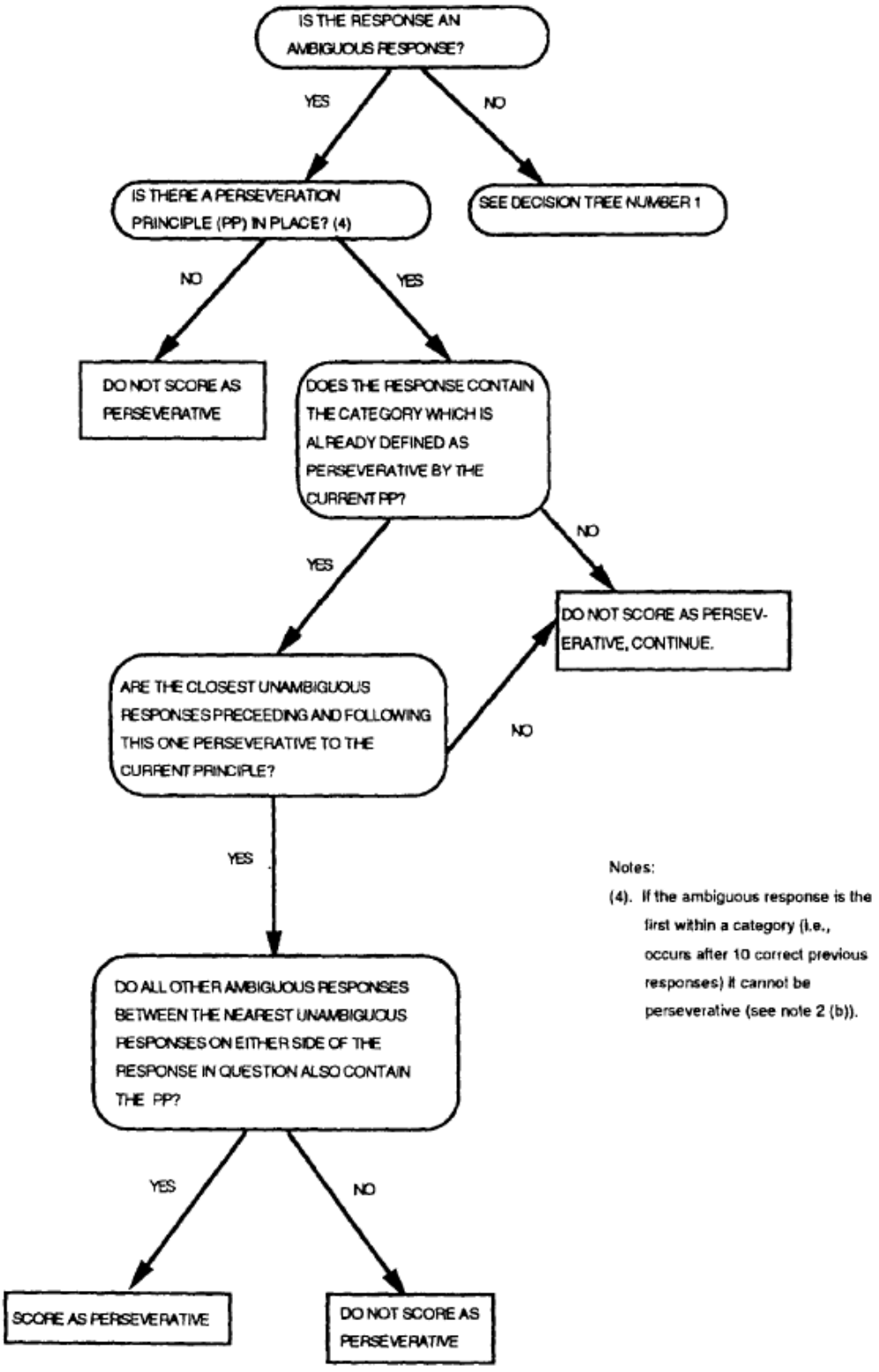


Figure 3. Steve Berry's original diagrammatic procedure for scoring ambiguous perseverative responses on the Wisconsin Card Sorting Test. Berry, S. Diagrammatic procedure for scoring the Wisconsin Card Sorting Test, *The Clinical Neuropsychologist*, © copyright [1996], reprinted with permission of Informa UK Limited, trading as Taylor & Taylor & Francis Group, http://www.tandfonline.com

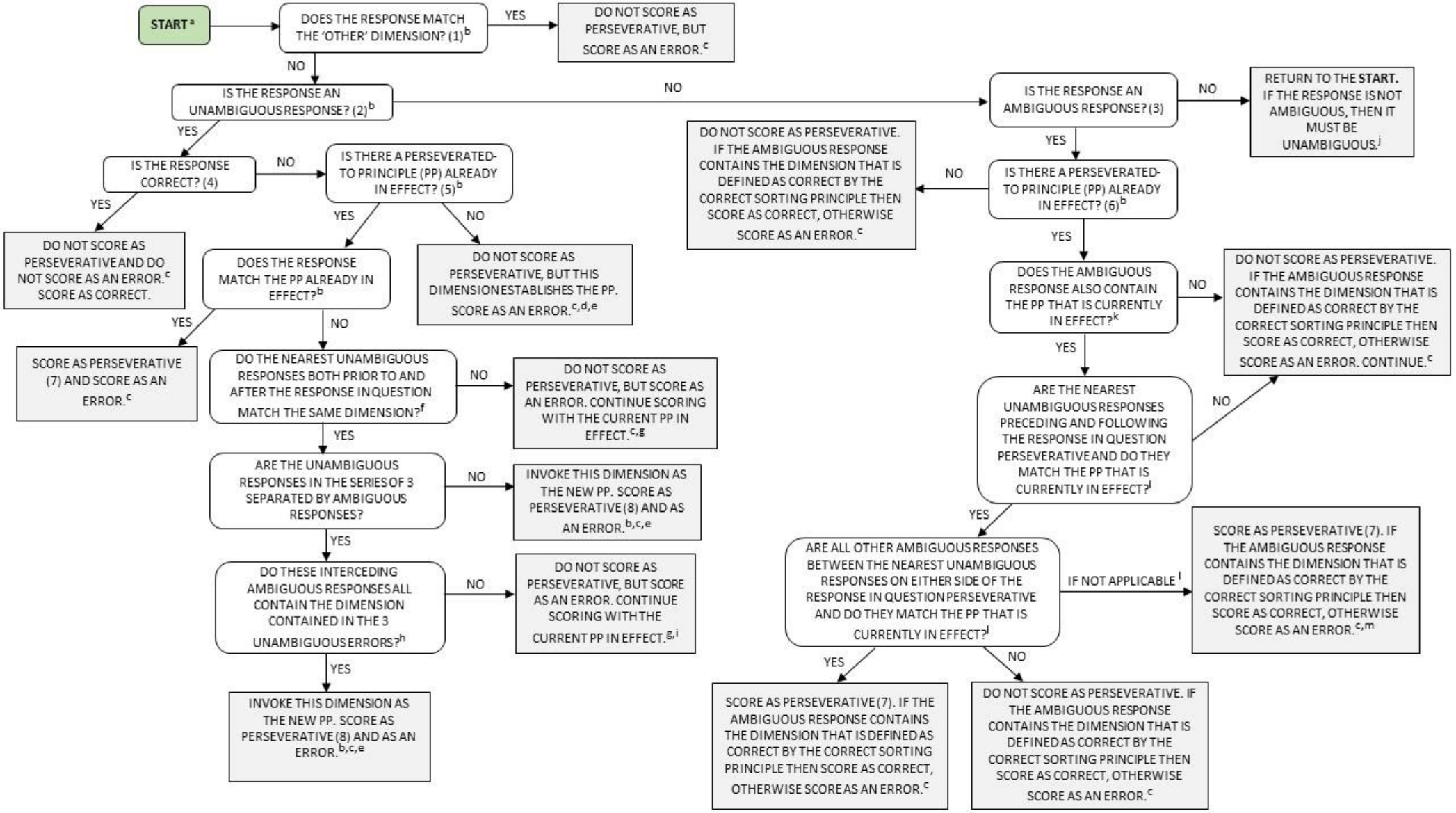


Figure 4. Adaptation of Steve Berry's 1996 diagrammatic procedure for scoring perseverative responses (and perseverative errors) on the Wisconsin Card Sorting Test

Corresponding notes:

1) A response that does not match either Colour, Form or Number (i.e., C F N Ø).
2) An unambiguous response is one in which the response card matches a stimulus card on only one dimension. For example, Ø F N O.
3) An ambiguous response is one in which the response card matches a stimulus card on more than one dimension. For example, Ø Ø N O. By definition, responses that match the 'Other' dimension are ambiguous.
4) A response is deemed to be *correct* when it matches the correct sorting principle that is currently in effect. If the response does not match the correct sorting principle that is currently in effect, the response is deemed to be incorrect and is scored as an error.
5) The perseverated-to principle (PP) is the rule that determines whether a response is considered *perseverative* or not. The PP is established in the first category (i.e., Colour category) once the first unambiguous error is made. The rule is determined by the examinee's responses and as depicted in the diagram, the PP can change throughout the task:
    i. immediately after a category is completed (i.e., the response after the 10th correct response), the PP is deemed to have lapsed. The PP becomes the dimension that was the correct sorting principle in the previous category
    ii. when 3 consecutive unambiguous errors are made in a row to a dimension that is not correct or perseverative, or
    iii. where there are ambiguous responses between these 3 unambiguous errors, they all match the same dimension that is neither correct, nor perseverative.
6) If an ambiguous response is made immediately after a category has been completed (i.e., occurs after 10 consecutive correct responses), it cannot be scored as perseverative.
7) Perseverative errors are perseverative responses; however, not all perseverative responses are perseverative errors. Perseverative responses can either be correct or incorrect as a result of ambiguous responses and the 'sandwich rule' (i.e., an ambiguous response that is situated ('sandwiched') between two unambiguous perseverative errors).
8) When new PP's are invoked, they do not come into effect, nor are they scored as perseverative until the second unambiguous error.[b,n]

This diagram has been adapted from Berry's (1996) three original diagrams, which have now been combined into a single diagram. A detailed summary of the changes is provided below; the footnotes correspond to the footnotes presented in the diagram (except for 'n') to indicate where changes have been made:

[a] 'Start' has been added to the diagram to indicate where users of the WCST should begin when scoring each individual response.

[b] Rephrased to use terminology that is consistent with the Heaton et al. (1993) WCST manual.

[c] Instructions for how to score perseverative errors have been incorporated into the diagram.

[d] This section of Berry's diagram was technically inaccurate. That is, if there is no PP already in effect, then the PP is not *new* (as Berry described), but is *established*. This section has been re-worded to accurately reflect that scenario.

[e] This section of Berry's diagram does not allow users of the WCST to score each response on an individual basis, so this has been adapted accordingly.

[f] 'Closest' has been replaced with 'nearest' for consistency.

[g] 'Previous PP' has been replaced with 'current PP in effect'. Using 'previous PP' implies that an earlier PP should be used, when in fact, the current PP remains in effect.

[h] This section of Berry's diagram is technically inaccurate. 'Unambiguous responses' has been replaced with 'unambiguous errors' because such responses also need to be incorrect (i.e., not matching the correct sorting principle).

[i] Explicit instructions have been provided for how to score this scenario because Berry (1996) did not provide such instructions (i.e., whether the response should be scored as perseverative or not).

[j] 'See decision tree number 1' has been replaced with 'return to start' and a brief description as to why users should return to the start of the diagram has been provided.

[k] This section has been simplified. Berry's original phrasing included both the term 'category' and 'PP', but introducing another term here is unnecessary.

[l] This section has been made clearer by stipulating that the unambiguous responses preceding and following the response in question must be i) perseverative and ii) match the PP that is currently in effect.

[m] This section has been added to the diagram so that in cases where no more than one ambiguous response is given, users of the WCST know how to score such scenarios (i.e., if not applicable, then score as perseverative).

[n] A more comprehensive notes section has been added that provides key information to consider when scoring perseverative responses and perseverative errors.

Finally, as outlined in Chapter 2, the WCST manual provides an overview of the key terminology and scoring dimensions (i.e., correct versus incorrect, ambiguous versus unambiguous, and perseverative versus non-perseverative) to help guide researchers and clinicians on how to score each response made by examinees. However, when it comes to describing other key aspects of the WCST, the wider literature has used varied terms that deviate from Heaton et al's. (1993) original descriptions. For instance, Heaton et al. (1993) referred to the individual characteristics by which the deck of cards could be sorted by (i.e., colour, form or number) as *'dimensions'*; however, previous studies have instead used terms such as *'attribute' 'category'* or *'criterion'* in place of *dimension*. Similarly, there remains confusion surrounding the differences between *perseverative responses* and *perseverative errors*, and as such these two terms have been used interchangeably. The fields' non-compliance with the specific terms and definitions of key variables outlined in the WCST manual has created vast opportunities for misunderstanding, and in turn, has impacted the interpretation of the scoring procedures and the analysis of key variables of the task (i.e., perseverative responses and perseverative errors). Potential solutions for rectifying these issues could be for Heaton et al. (1993) to i) consider incorporating a table with various terms and definitions that are relevant to the administration and scoring of the task, and ii) make clearer the overlap and distinction between perseverative responses and perseverative errors in the WCST manual by including visual illustrations – similar to that of Miles et al., (2021).

7.7 Concluding remarks

Overall, this thesis investigated the potential role of cognitive flexibility – as assessed via self-report and neuropsychological assessments – in persistent pain conditions. In doing so, important discoveries were made, which required substantial adjustments to the original PhD plan. The key findings from the studies contained in this thesis were that implementing different scoring procedures of the WCST has vast ramifications when comparing findings across studies; cognitive flexibility is a poorly understood and defined construct; self-report and neuropsychological assessments of cognitive flexibility are often conflated and used interchangeably, despite the fact that empirical evidence points to a lack of relationship between these two approaches; and that once important confounders are controlled for, people with persistent pain appear to be no less capable than pain-free controls in adapting their cognitive and behavioural strategies when faced with changing task or environmental demands. These findings have important theoretical and practical implications for the field, and offer possible avenues for future research to address the

lack of consensus and ambiguity surrounding cognitive flexibility, and how to improve the practical use, and clinical relevance, of cognitive flexibility assessments moving forward.

7.8 References

Ahern, M., Dean, C. M., Dear, B. F., Willcock, S. M., & Hush, J. M. (2019). The experiences and needs of people seeking primary care for low-back pain in Australia. *Pain Reports*, *4*(4), e756. https://doi.org/10.1136/bjsm.2007.037747

Alessi, E. J., & Martin, J. I. (2010). Conducting an Internet-based survey: Benefits, pitfalls, and lessons learned. *Social Work Research*, *34*(2), 122–128. https://doi.org/10.1093/swr/34.2.122

Alyas, F., Turner, M., & Connell, D. (2007). MRI findings in the lumbar spines of asymptomatic, adolescent, elite tennis players. *British Journal of Sports Medicine*, *41*(11), 836–841. https://doi.org/10.1136/bjsm.2007.037747

Anand, K. J. S., & Craig, K. D. (1996). New perspectives on the definition of pain. *Pain*, *67*(1), 3–6. https://doi.org/10.1016/0304-3959(96)03135-1

Arnow, B. A., Hunkeler, E. M., Blasey, C. M., Lee, J., Constantino, M. J., Fireman, B., … Hayward, C. (2006). Comorbid depression, chronic pain, and disability in primary care. *Psychosomatic Medicine*, *68*(2), 262–268. https://doi.org/10.1097/01.psy.0000204851.15499.fc

Arntz, A., & Claassens, L. (2004). The meaning of pain influences its experienced intensity. *Pain*, *109*(1), 20–25. https://doi.org/10.1016/j.pain.2003.12.030

Asali, A., Miller, N., Engel, O., Haj-Yahia, N., Cohen, O., Ben-Kimhy, R., & Biron-Shental, T. (2020). Does cognitive flexibility affect the course of labor? A prospective study. *International Journal of Gynecology & Obstetrics*, *148*(2), 187–191. https://doi.org/10.1002/ijgo.13063

Ashendorf, L., & McCaffrey, R. J. (2008). Exploring age-related decline on the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *22*(2), 262–272. https://doi.org/10.1080/13854040701218436

Attal, N., Masselin-Dubois, A., Martinez, V., Jayr, C., Albi, A., Fermanian, J., … Baudic, S. (2014). Does cognitive functioning predict chronic pain? Results from a prospective surgical cohort. *Brain*, *137*(3), 904–917. https://doi.org/10.1093/brain/awt354

Axelrod, B. N., Goldman, R. S., & Woodard, J. L. (1992). Interrater reliability in scoring the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *6*(2), 143–155. https://doi.org/10.1080/13854049208401851

Aydede, M. (2019). Does the IASP definition of pain need updating? *Pain Reports*, *4*(5), e777.

https://doi.org/10.1097/PR9.0000000000000777

Badre, D., & Wagner, A. D. (2006). Computational and neurobiological mechanisms underlying cognitive flexibility. *Proceedings of the National Academy of Sciences*, *103*(18), 7186–7191. https://doi.org/10.1073/pnas.0509550103

Baker, K. S., Georgiou-Karistianis, N., Lampit, A., Valenzuela, M., Gibson, S. J., & Giummarra, M. J. (2018). Computerised training improves cognitive performance in chronic pain: A participant-blinded randomised active-controlled trial with remote supervision. *Pain*, *159*(4), 644–655. https://doi.org/10.1097/j.pain.0000000000001150

Baker, K. S., Gibson, S., Georgiou-Karistianis, N., Roth, R. M., & Giummarra, M. J. (2016). Everyday executive functioning in chronic pain: Specific deficits in working memory and emotion control, predicted by mood, medications, and pain interference. *The Clinical Journal of Pain*, *32*(8), 673–680. https://doi.org/10.1097/AJP.0000000000000313

Ball, H. L. (2019). Conducting online surveys. *Journal of Human Lactation*, *35*(3), 413–417. https://doi.org/10.1177/0890334419848734

Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, *13*(3), 283–292. https://doi.org/10.1076/clin.13.3.283.1743

Basso, M. R., Lowery, N., Ghormley, C., & Bornstein, R. A. (2001). Practice effects on the Wisconsin Card Sorting Test-64 card version across 12 months. *The Clinical Neuropsychologist*, *15*(4), 471–478. https://doi.org/10.1076/clin.15.4.471.1883

Bayer, T. L., Baer, P. E., & Early, C. (1991). Situational and psychophysiological factors in psychologically induced pain. *Pain*, *44*(1), 45–50. https://doi.org/10.1016/0304-3959(91)90145-N

Bell, T., Trost, Z., Buelow, M. T., Clay, O., Younger, J., Moore, D., & Crowe, M. (2018). Meta-analysis of cognitive performance in fibromyalgia. *Journal of Clinical and Experimental Neuropsychology*, *40*(7), 698–714. https://doi.org/10.1080/13803395.2017.1422699

Ben-Zion, Z., Fine, N. B., Keynan, N. J., Admon, R., Green, N., Halevi, M., … Shalev, A. Y. (2018). Cognitive flexibility predicts PTSD symptoms: Observational and interventional studies. *Frontiers in Psychiatry*, *9*, 477. https://doi.org/10.3389/fpsyt.2018.00477

Benton, A. L., Hamsher, K. D., & Sivan, A. B. (1994). Controlled Oral Word Association Test.

Multilingual Aphasia Examination, *3*.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, *39*(1), 15–22. https://doi.org/10.1080/00221309.1948.9918159

Bergman, S., Herrstrom, P., Hogstrom, K., Petersson, I. F., Svensson, B., & Jacobsson. L T H. (2001). Chronic musculoskeletal pain, prevalence rates, and sociodemographic associations in a Swedish population study. *The Journal of Rheumatology*, *28*(6), 1369–1377.

Berry, S. (1996). Diagrammatic procedure for scoring the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *10*(1), 117–121. https://doi.org/10.1080/13854049608406670

Berryman, C., Stanton, T. R., Bowering, K. J., Tabor, A., McFarlane, A., & Moseley, G. L. (2014). Do people with chronic pain have impaired executive function? A meta-analytical review. *Clinical Psychology Review*, *34*(7), 563–579. https://doi.org/10.1016/j.cpr.2014.08.003

Bevers, K., Watts, L., Kishino, N. D., & Gatchel, R. J. (2016). The biopsychosocial model of the assessment, prevention, and treatment of chronic pain. *US Neurology*, *12*(2), 98–104. https://doi.org/10.17925/USN.2016.12.02.98

Bezuglov, E., Lazarev, A., Petrov, A., Brodskaia, A., Lyubushkina, A., Kubacheva, K., … Nikolenko, V. (2021). Asyptomatic degenerative changes in the lumbar spine among professional soccer players. *Spine*, *46*(2), 122–128. https://doi.org/10.1097/BRS.0000000000003726

Bliss, T. V. P., & Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, *361*, 31–39. https://doi.org/10.1038/361031a0

Bliss, T. V. P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology*, *232*(2), 331–356. https://doi.org/10.1113/jphysiol.1973.sp010273

Blyth, F. M., March, L. M., Brnabic, A. J. M., & Cousins, M. J. (2004). Chronic pain and frequenct use of health care. *Pain*, *111*(1), 51–58. https://doi.org/10.1016/j.pain.2004.05.020

Blyth, F. M., March, L. M., Brnabic, A. J. M., Jorm, L. R., Williamson, M., & Cousins, M. J. (2001). Chronic pain in Australia: A prevalence study. *Pain*, *89*(2), 127–134. https://doi.org/10.1016/S0304-3959(00)00355-9

Bondesson, E., Larrosa Pardo, F., Stigmar, K., Ringqvist, A., Petersson, I. F., Joud, A., & Schelin, M. E. C. (2018). Comorbidity between pain and mental illness - Evidence of a bidirectional

relationship. *European Journal of Pain*, *22*(7), 1304–1311. https://doi.org/10.1002/ejp.1218

Borenstein, M. (2022). Comprehensive Meta-Analysis software. In Systematic Reviews in Health Research (Eds. M Egger, J. P. T. Higgins. and G. Davey Smith). https://doi.org/10.1002/9781119099369.ch27

Braem, S. (2017). Conditioning task switching behavior. *Cognition*, *166*, 272–276. https://doi.org/10.1016/j.cognition.2017.05.037

Braem, S., & Egner, T. (2018). Getting a grip on cognitive flexibility. *Current Directions in Psychological Science*, *27*(6), 470–476. https://doi.org/10.1177/0963721418787475

Breivik, H., Collett, B., Ventafridda, V., Cohen, R., & Gallacher, D. (2006). Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. *European Journal of Pain*, *10*(4), 287–333. https://doi.org/10.1016/j.ejpain.2005.06.009

Büchel, C., Geuter, S., Sprenger, C., & Eippert, F. (2014). Placebo analgesia: A predictive coding perspective. *Neuron*, *81*(6), 1223–1239. https://doi.org/10.1016/j.neuron.2014.02.042

Bunk, S., Preis, L., Zuidema, S., Lautenbacher, S., & Kunz, M. (2019). Executive functions and pain. *Zeitschrift Für Neuropsychologie*, *30*(3), 169–196. https://doi.org/10.1024/1016-264X/a000264

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton tests*.

Butler, D. S., & Moseley, G. L. (2013). *Explain Pain* (2nd ed.). Adelaide: Noigroup Publications.

Buttelmann, F., & Karbach, J. (2017). Development and plasticity of cognitive flexibility in early and middle childhood. *Frontiers in Psychology*, *8*, 1040. https://doi.org/10.3389/fpsyg.2017.01040

Campbell, C. M., & Edwards, R. R. (2012). Ethnic differences in pain and pain management. *Pain Management*, *2*(3), 219–230. https://doi.org/10.2217/pmt.12.7

Cañas, J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics*, *46*(5), 482–501. https://doi.org/10.1080/0014013031000061640

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. M. (2001). Changes in executive control across the lifespan: examination of task-switching performance. *Developmental Psychology*, *37*(5), 715–730. https://doi.org/10.1037/0012-1649.37.5.715

Chalmers, K. J. (2022). Commentary on “The role of psychological flexibility, perceived injustice,

and body-image in vulvodynia: A longitudinal study" by Chisari et al. *European Journal of Pain*, *26*(1), 5–6. https://doi.org/10.1002/ejp.1857

Chaves, T. C., Stanton, T. R., Grant, A., Pulling, B. W., Madden, V. J., Newport, R., & Moseley, G. L. (2021). Imprecise visual feedback about hand location increases a classically conditioned pain expectancy effect. *The Journal of Pain*, *22*(6), 748–761. https://doi.org/10.1016/j.jpain.2021.01.004

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*, 181–197. https://doi.org/10.1023/B:NERV.0000009483.91468.fb

Chen, L., Ferreira, M. L., Nassar, N., Preen, D. B., Hopper, J. L., Li, S., … Ferreira, P. H. (2021). Association of chronic musculoskeletal pain with mortality among UK adults: A population-based cohort study with mediation analysis. *EClinicalMedicine*, *42*, 101202. https://doi.org/10.1016/j.eclinm.2021.101202

Cherry, B. J., Zettel-Watson, L., Shimizu, R., Roberson, I., Rutledge, D. N., & Jones, C. J. (2014). Cognitive performance in women aged 50 years and older with and without fibromyalgia. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *69*(2), 199–208. https://doi.org/10.1093/geronb/gbs122

Cherry, K. M., Vander Hoeven, E., Patterson, T. S., & Lumley, M. N. (2021). Defining and measuring "psychological flexibility": A narrative scoping review of diverse flexibility and rigidity constructs and perspectives. *Clinical Psychology Review*, *84*, 101973. https://doi.org/10.1016/j.cpr.2021.101973

Chicchi Giglioli, I. A., Gálvez, B. P., Granados, A. G., & Raya, M. A. (2021). The virtual cooking task: A preliminary comparison between neuropsychological and ecological virtual reality tests to assess executive functions alterations in patients affected by alcohol use disorder. *Cyberpsychology, Behavior, and Social Networking*, *24*(10), 673–682. https://doi.org/10.1089/cyber.2020.0560

Chisari, C., Budhraja, M., Monajemi, M. B., Lewis, F., Moss-Morris, R., Scott, W., & McCracken, L. M. (2022). The role of psychological flexibility, perceived injustice and body image in vulvodynia: A longitudinal study. *European Journal of Pain*, *26*(1), 103–113. https://doi.org/10.1002/ejp.1841

Clark, A. (2015). *Surfing uncertainty: prediction, action and the embodied mind*. Oxford University Press.

Clark, R. E. (2004). The classical origins of Pavlov's conditioning. *Integrative Physiological & Behavioral Science*, *39*(4), 279–294.

Colé, P., Duncan, L. G., & Blaye, A. (2014). Cognitive flexibility predicts early reading skills. *Frontiers in Psychology*, *5*, 565. https://doi.org/10.3389/fpsyg.2014.00565

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., … Schulte-Mecklenbeck, M. (2019). 7 easy steps to open science: An annotated reading list. *Zeitschrift Für Psychologie*, *227*(4), 237–248. https://doi.org/10.1027/2151-2604/a000387

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571–578. https://doi.org/10.1016/j.tins.2015.07.003

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences, 24*(4), 267-269. https://doi.org/10.1016/j.tics.2020.01.007

Davis, J. C., Marra, C. A., Najafzadeh, M., & Liu-Ambrose, T. (2010). The independent contribution of executive functions to health related quality of life in older women. *BMC Geriatrics*, *10*(16), 1–8. https://doi.org/10.1186/1471-2318-10-16

Deák, G. O. (2003). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior*, *31*, 271–327. https://doi.org/10.1016/S0065-2407(03)31007-9

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). Delis-Kaplan Executive Functioning System. *Assessment*.

Deloitte Access Economics. (2019). The cost of pain in Australia. Retrieved August 25, 2021, from https://www2.deloitte.com/au/en/pages/economics/articles/cost-pain-australia.html

Demirtas, A. S., & Yildiz, B. (2019). Hopelessness and perceived stress: The mediating role of cognitive flexibility and intolerance of uncertainty. *Dusunen Adam The Journal of Psychiatry and Neurological Sciences*, *3*(32), 259–267. https://doi.org/10.14744/DAJPNS.2019.00035

Dennis, J. P., & Vander Wal, J. S. (2010). The Cognitive Flexibility Inventory: Instrument

development and estimates of reliability and validity. *Cognitive Therapy and Research*, *34*(3), 241–253. https://doi.org/10.1007/s10608-009-9276-4

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dimsdale, J. E., & Dantzer, R. (2007). A biological substrate for somatoform disorders: Importance of pathophysiology. *Psychosomatic Medicine*, *69*(9), 850–854. https://doi.org/10.1097/PSY.0b013e31815b00e7

Doorley, J. D., Goodman, F. R., Kelso, K. C., & Kashdan, T. B. (2020). Psychological flexibility: What we know, what we do not know, and what we think we know. *Social and Personality Psychology Compass*, *14*(12), 1–11. https://doi.org/10.1111/spc3.12566

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ Open*, *6*(12), e011458. https://doi.org/10.1136/bmjopen-2016-011458

Drapeau, C. E., & Bastien-Toniazzo, M. (2007). Nonequivalence of computerized and paper-and-pencil versions of the Trail Making Test. *Perceptual and Motor Skills2*, *104*, 785–791. https://doi.org/10.2466/pms.104.3.785-791

El-Den, S., Schneider, C., Mirzaei, A., & Carter, S. (2020). How to measure a latent construct: Psychometric principles for the development and validation of measurement instruments. *International Journal of Pharmacy Practice*, *28*(4), 326–336. https://doi.org/10.1111/ijpp.12600

Eliava, N. (1964). *A problem of set in cognitive psychology*. Tbilisi, Georgia: Academic Press.

Eling, P., Derckx, K., & Maes, R. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, *67*(3), 247–253. https://doi.org/10.1016/j.bandc.2008.01.006

Elliott, A. M., Smith, B. H., Hannaford, P. C., Smith, W. C., & Chambers, W. A. (2002). The course of chronic pain in the community: Results of a 4-year follow-up study. *Pain*, *99*(1), 299–307. https://doi.org/10.1016/S0304-3959(02)00138-0

Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, *196*(4286), 129–136. https://doi.org/10.1126/science.847460

Eslinger, P. J., & Gratten, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia*, *31*(1), 17–28. https://doi.org/10.1016/0028-3932(93)90077-D

Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research*, *6*(3), e34. https://doi.org/10.2196/jmir.6.3.e34

Faucett, J., Gordon, N., & Levine, J. (1994). Differences in postoperative pain severity among four ethnic groups. *Journal of Pain and Symptom Management*, *9*(6), 383–389. https://doi.org/10.1016/0885-3924(94)90175-9

Fayaz, A., Croft, P., Langford, R. M., Donaldson, L. J., & Jones, G. T. (2016). Prevalence of chronic pain in the UK: A systematic review and meta-analysis of population studies. *BMJ Open*, *6*(6), e010364. https://doi.org/10.1136/bmjopen-2015-010364

Fisher, J. P., Hassan, D. T., & O'Connor, N. (1995). Minerva (Emergency Department Case Study). *BMJ*, *310*, 70. https://doi.org/10.1136/bmj.310.6971.70

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Flashman, L. A., Horner, M. D., & Freides, D. (1991). Note on scoring perseveration on the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *5*(2), 190–194. https://doi.org/10.1080/13854049108403303

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window of cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of The Royal Society*, *360*, 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138. https://doi.org/10.1038/nrn2787

Fujino, J., Tei, S., Itahashi, T., Aoki, Y., Ohta, H., Kubota, M., … Takahashi, H. (2019). Need for closure and cognitive flexibility in individuals with autism spectrum disorder: A preliminary study. *Psychiatry Research*, *271*, 247–252. https://doi.org/10.1016/j.psychres.2018.11.057

Garcia-Willingham, N. E., Roach, A. R., Kasarskis, E. J., & Segerstrom, S. C. (2018). Self-regulation and executive functioning as related to survival in Motor Neuron Disease: Preliminary findings. *Psychosomatic Medicine*, *80*(7), 665–672. https://doi.org/10.1097/PSY.0000000000000602

Gaskin, D. J., & Richard, P. (2012). The economic costs of pain in the United States. *The Journal of Pain*, *13*(8), 715–724. https://doi.org/10.1016/j.jpain.2012.03.009

Gatchel, R. J., McGeary, D. D., McGeary, C. A., & Lippe, B. (2014). Interdisciplinary chronic pain management: Past, present, and future. *American Psychologist*, *69*(2), 119–130. https://doi.org/10.1037/a0035514

Gatchel, R. J., Peng, Y. B., Peters, M. L., Fuchs, P. N., & Turk, D. C. (2007). The biopsychosocial approach to chronic pain: Scientific advances and future directions. *Psychological Bulletin*, *133*(4), 581–624. https://doi.org/10.1037/0033-2909.133.4.581

Geiser, D. S. (1992). *A comparison of acceptance-focused and control-focused psychological treatments in a chronic pain treatment center*. University of Nevada-Reno.

Gelonch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and objective measures. *Comprehensive Psychiatry*, *66*, 113–122. https://doi.org/10.1016/j.comppsych.2016.01.002

Gentili, C., Rickardsson, J., Zetterqvist, V., Simons, L. E., Lekander, M., & Wicksell, R. K. (2019). Psychological flexibility as a resilience factor in individuals with chronic pain. *Frontiers in Psychology*, *10*, 2016. https://doi.org/10.3389/fpsyg.2019.02016

Geurts, H. M., Corbett, B., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in Cognitive Sciences*, *13*(2), 74–82. https://doi.org/10.1016/j.tics.2008.11.006

Gonzalez, C. A., Figueroa, I. J., Bellows, B. G., Rhodes, D., & Youmans, R. J. (2013). A new behavioral measure of cognitive flexibility. *In: Harris, D. (Eds) Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition. EPCE 2013. Lecture Notes in Computer Science, Vol 8019*, 297–306. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-39360-0_33

Gosling, S. D., Sandy, C. J., John, O. P., & Potter, J. (2010). Wired but not WEIRD: The promise of the Internet in reaching more diverse samples. *Behavioral and Brain Sciences*, *33*(2–3), 94–95. https://doi.org/10.1017/S0140525X10000300

Grant, A., & Cassidy, S. (2022). Exploring the relationship between psychological flexibility and self-report and task-based measures of cognitive flexibility. *Journal of Contextual Behavioral Science*, *23*, 144–150. https://doi.org/10.1016/j.jcbs.2021.12.006

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Greve, K. W. (1993). Can perseverative responses on the Wisconsin Card Sorting Test be scored accurately? *Archives of Clinical Neuropsychology*, *8*(6), 511–517. https://doi.org/10.1016/0887-6177(93)90051-2

Griffin, J. W. (2021). Calculating statistical power for meta-analysis using metapower. *The Quantitative Methods for Psychology*, *17*(1), 24–39. https://doi.org/10.20982/tqmp.17.1.p024

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, *53*(4), 267–293. https://doi.org/10.1037/h0040755

Hadi, M. A., McHugh, G. A., & Closs, S. J. (2019). Impact of chronic pain on patients' quality of life: A comparative mixed-methods study. *Journal of Patient Experience*, *6*(2), 133–141. https://doi.org/10.1177/2374373518786013

Harvie, D. S., Broecker, M., Smith, R. T., Meulders, A., Madden, V. J., & Moseley, G. L. (2015). Bogus visual feedback alters onset of movement-evoked pain in people with neck pain. *Psychological Science*, *26*(4), 385–392. https://doi.org/10.1177/0956797614563339

Harvie, D. S., Meulders, A., Madden, V. J., Hillier, S. L., Peto, D. K., Brinkworth, R., & Moseley, G. L. (2016). When touch predicts pain: predictive tactile cues modulate perceived intensity of painful stimulation independent of expectancy. *Scandinavian Journal of Pain*, *11*(1), 11–18. https://doi.org/10.1016/j.sjpain.2015.09.007

Harvie, D. S., Meulders, A., Reid, E., Camfferman, D., Brinkworth, R. S. A., & Moseley, G. L. (2016). Selectivity of conditioned fear of touch is modulated by somatosensory precision. *Psychophysiology*, *53*(6), 921–929. https://doi.org/10.1111/psyp.12631

Harvie, D. S., Moseley, G. L., Hillier, S. L., & Meulders, A. (2017). Classical conditioning differences

associated with chronic pain: A systematic review. *The Journal of Pain*, *18*(8), 889–898. https://doi.org/10.1016/j.jpain.2017.02.430

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, *44*(1), 1–25. https://doi.org/10.1016/j.brat.2005.06.006

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: Understanding and treating human suffering.* New York: Guilford Press.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual*. Psychological Assessment Resources.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Psychological Assessment Resources.

Hebb, D. O. (1949). *The organization of behaviour: A neuropsychology theory*. New York: Wiley.

Henningsen, P., Gündel, H., Kop, W., Löwe, B., Martin, A., Rief, W., … Van den Bergh, O. (2018). Persistent physical symptoms as perceptual dysregulation: A neuropsychobehavioral model and its clinical implications. *Psychosomatic Medicine*, *80*(5), 422–431. https://doi.org/10.1097/PSY.0000000000000588

Herbrich, L. R., Kappel, V., Winter, S. M., & van Noort, B. M. (2019). Executive functioning in adolescent anorexia nervosa: Neuropsychology versus self- and parental-report. *Child Neuropsychology*, *25*(6), 816–825. https://doi.org/10.1080/09297049.2018.1536200

Hohwy, J. (2013). *The predictive mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, *35*(2), 209–223. https://doi.org/10.1111/mila.12281

Howlett, C. A., Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., … Moseley, G. L. (2021). Same room - different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, *88*, 102061. https://doi.org/10.1016/j.cpr.2021.102061

Hoy, D., Bain, C., Williams, G., March, L., Brooks, P., Blyth, F., … Buchbinder, R. (2012). A systematic review of the global prevalence of low back pain. *Arthritis and Rheumatism*, *64*(6), 2028–2037.

https://doi.org/10.1002/art.34347

Hruschak, V., & Cochran, G. (2018). Psychosocial predictors in the transition from acute to chronic pain: A systematic review. *Psychology, Health & Medicine*, *23*(10), 1151–1167. https://doi.org/10.1080/13548506.2018.1446097

IASP Subcommittee on Taxonomy, (1979). The need of a taxonomy. *Pain*, *6*(3), 247–252. https://doi.org/10.1016/0304-3959(79)90046-0

Igumbor, E. U., Puoane, T., Gansky, S. A., & Plesh, O. (2012). Pain as a reason for primary care visits: Cross-sectional survey in a rural and periurban health clinic in the Eastern Cape, South Africa. *South African Family Practice*, *54*(3), 256–263. https://doi.org/10.1080/20786204.2012.10874224

Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, *30*(2), 190–200. https://doi.org/10.1016/j.newideapsych.2011.11.001

Ionescu, T. (2017). The variability-stability-flexibility pattern: A possible key to understanding the flexibility of the human mind. *Review of General Psychology*, *21*(2), 123–131. https://doi.org/10.1037/gpr0000110

Isquith, P. K., Roth, R. M., & Gioia, G. (2013). Contribution of rating scales to the assessment of executive functions. *Applied Neuropsychology: Child*, *2*(2), 125–132. https://doi.org/10.1080/21622965.2013.748389

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., … Murray, C. J. L. (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, *392*(10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7

Janevic, M. R., McLaughlin, S. J., Heapy, A. A., Thacker, C., & Piette, J. D. (2017). Racial and socioeconomic disparities in disabling chornic pain: Findings from the health and retirement study. *The Journal of Pain*, *18*(12), 1459–1467. https://doi.org/10.1016/j.jpain.2017.07.005

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). Reliability and validity of two self-report measures of cognitive flexibility. *Psychological Assessment*, *26*(4), 1381–1387. https://doi.org/10.1037/a0038009

Jonas, W. B., Crawford, C., Colloca, L., Kriston, L., Linde, K., Moseley, B., & Meissner, K. (2019). Are

invasive procedures effective for chronic pain? A systematic review. *Pain Medicine*, *20*(7), 1281–1293. https://doi.org/10.1093/pm/pny154

Kaptchuk, T. J., Hemond, C. C., & Miller, F. G. (2020). Placebos in chronic pain: Evidence, theory, ethics and use in clinical practice. *BMJ*, *370*, m1668. https://doi.org/10.1136/bmj.m1668

Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. *Collabra: Psychology*, *7*(1), 18684. https://doi.org/10.1525/collabra.18684

Kawai, K., Kawai, A. T., Wollan, P., & Yawn, B. P. (2017). Adverse impacts of chronic pain on health-related quality of life, work productivity, depression and anxiety in a community-based study. *Family Practice*, *34*(6), 656–661. https://doi.org/10.1093/fampra/cmx034

Kercood, S., Lineweaver, T. T., Frank, C. C., & Fromm, E. D. (2017). Cognitive flexibility and its relationship to academic achievement and career choice of college students with and without attention deficit hyperactivity disorder. *Journal of Postsecondary Education and Disability*, *30*(4), 329–344.

Kessels, R. P. C., Aleman, A., Verhagen, W. I. M., & van Luijtelaar, E. L. J. M. (2000). Cognitive functioning after whiplash injury: A meta-analysis. *Journal of the International Neuropsychological Society*, *6*(3), 271–278. https://doi.org/10.1017/S1355617700633027

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., … Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1), 20. https://doi.org/10.1525/collabra.158

Knill, D. C., & Pouget, A. (2004). The Bayseian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kopp, B., Rösser, N., & Wessel, K. (2008). Psychometric characteristics and practice effects of the Brunswick Trail Making Test. *Perceptual and Motor Skills*, *107*(3), 707–733. https://doi.org/10.2466/PMS.107.3.707-733

Kupis, L., Goodman, Z. T., Kornfeld, S., Hoang, S., Romero, C., Dirks, B., … Uddin, L. Q. (2021). Brain dynamics underlying cognitive flexibility across the lifespan. *Cerebral Cortex*, *31*(11), 5263–5274. https://doi.org/10.1093/cercor/bhab156

Lange, F. (2020). Are difficult-to-study populations too difficult to study in a reliable way?: Lessons

learned from meta-analyses in clinical neuropsychology. *European Psychologist*, *25*(1), 41–50. https://doi.org/10.1027/1016-9040/a000384

Latremoliere, A., & Woolf, C. J. (2009). Central sensitization: A generator of pain hypersensitivity by central neural plasticity. *The Journal of Pain*, *10*(9), 895–926. https://doi.org/10.1016/j.jpain.2009.06.012

Lawson, K. M., & Robins, R. W. (2021). Sibling constructs: What are they, why do they matter, and how should you handle them? *Personality and Social Psychology Review*, *25*(4), 344–366. https://doi.org/10.1177/10888683211047101

Lee, D.-H., Lee, K.-J., Cho, K. I. K., Noh, E. C., Jang, J. H., Kim, Y. C., & Kang, D.-H. (2015). Brain alterations and neurocognitive dysfunction in patients with complex regional pain syndrome. *The Journal of Pain*, *16*(6), 580–586. https://doi.org/10.1016/j.jpain.2015.03.006

Lee, H., Lamb, S. E., Bagg, M. K., Toomey, E., Cashin, A. G., & Moseley, G. L. (2018). Reproducible and replicable pain research: A critical review. *Pain*, *159*(9), 1683–1689. https://doi.org/10.1097/j.pain.0000000000001254

Lee, R. S. C., Hermens, D. F., Porter, M. A., & Redoblado-Hodge, M. A. (2012). A meta-analysis of cognitive deficits in first-episode major depressive disorder. *Journal of Affective Disorders*, *140*(2), 113–124. https://doi.org/10.1016/j.jad.2011.10.023

Lewis, J., & O'Sullivan, P. (2018). Is it time to reframe how we care for people with non-traumatic musculoskeletal pain? *British Journal of Sports Medicine*, *52*(24), 1543–1544. https://doi.org/10.1136/bjsports-2018-099198

Lin, I., Wiles, L., Waller, R., Goucke, R., Nagree, Y., Gibberd, M., … O'Sullivan, P. P. B. (2020). What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: Systematic review. *British Journal of Sports Medicine*, *54*(2), 79–86. https://doi.org/10.1136/bjsports-2018-099878

Loeser, J. D. (1982). Concepts of pain. In *M. Stanton-Hicks & R. Boaz (Eds.), Chronic low back pain (pp. 145-148)*. New York, US: Raven Press.

Loeser, J. D. (2000). Pain and suffering. *The Clinical Journal of Pain*, *16*(2), S2–S6. https://doi.org/10.1097/00002508-200006001-00002

Lounes, N., Khan, G., & Tchanturia, K. (2011). Assessment of cognitive flexibility in anorexia nervosa - self-report or experimental measure? A brief report. *Journal of the International*

Neuropsychological Society, *17*(5), 925–928. https://doi.org/10.1017/S1355617711000671

Louw, A., Diener, I., Fernandez-de-las-Penas, C., & Puentedura, E. J. (2017). Sham surgery in orthopedics: A systematic review of the literature. *Pain Medicine*, *18*(4), 736–750. https://doi.org/10.1093/pm/pnw164

Louw, A., Schmidt, S., Zimney, K., & Puentedura, E. J. (2019). Treat the patient, not the label: A pain neuroscience update. *Journal of Women's Health Physical Therapy*, *43*(2), 89–97. https://doi.org/10.1097/JWH.0000000000000121

Macfarlane, G. J., Beasley, M., Smith, B. H., Jones, G. T., & Macfarlane, T. V. (2015). Can large surveys conducted on highly selected populations provide valid information on the epidemiology of common health conditions? An analysis of UK Biobank data on musculoskeletal pain. *British Journal of Pain*, *9*(4), 203–212. https://doi.org/10.1177/2049463715569806

Madden, V. J., Bellan, V., Russek, L. N., Camfferman, D., Vlaeyen, J. W. S., & Moseley, G. L. (2016). Pain by association? Experimental modulation of human pain thresholds using classical conditioning. *The Journal of Pain*, *17*(10), 1105–1115. https://doi.org/10.1016/j.jpain.2016.06.012

Madden, V. J., Harvie, D. S., Parker, R., Jensen, K. B., Vlaeyen, J. W. S., Moseley, G. L., & Stanton, T. R. (2015). Can pain or hyperalgesia be a classically conditioned response in humans? A systematic review and meta-analysis. *Pain Medicine*, *17*(6), 1094–1111. https://doi.org/10.1093/pm/pnv044

Madden, V. J., & Moseley, G. L. (2016). Do clinicians think that pain can be a classically conditioned response to a non-noxious stimulus? *Manual Therapy*, *22*, 165–173. https://doi.org/10.1016/j.math.2015.12.003

Madden, V. J., Russek, L. N., Harvie, D. S., Vlaeyen, J. W. S., & Moseley, G. L. (2017). Classical conditioning fails to elicit allodynia in an experimental study with healthy humans. *Pain Medicine*, *18*(7), 1314–1325. https://doi.org/10.1093/pm/pnw221

Mäntyselkä, P., Kumpusalo, E., Ahonen, R., Kumpusalo, A., Kauhanen, J., Viinamäki, H., … Takala, J. (2001). Pain as a reason to visit the doctor: A study in Finnish primary health care. *Pain*, *89*(2), 175–180. https://doi.org/10.1016/S0304-3959(00)00361-4

Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports*,

76(2), 623–626. https://doi.org/10.2466/pr0.1995.76.2.623

Martin, M. M., & Anderson, C. M. (1998). The cognitive flexibility scale: Three validity studies. *Communication Reports*, *11*(1), 1–9. https://doi.org/10.1080/08934219809367680

Martucci, K. T., & Mackey, S. C. (2018). Neuroimaging of pain: Human evidence and clinical relevance of central nervous system processes and modulation. *Anesthesiology*, *128*(6), 1241–1254. https://doi.org/10.1097/ALN.0000000000002137

McCarberg, B., & Peppin, J. (2019). Pain pathways and nervous system plasticity: Learning and memory in pain. *Pain Medicine*, *20*(12), 2421–2437. https://doi.org/10.1093/pm/pnz017

McCracken, L. M., & Gutiérrez-Martínez, O. (2011). Processes of change in psychological flexibility in an interdisciplinary group-based treatment for chronic pain based on acceptance and commitment therapy. *Behaviour Research and Therapy*, *49*(4), 267–274. https://doi.org/10.1016/j.brat.2011.02.004

McCracken, L. M., Vowles, K. E., & Eccleston, C. (2004). Acceptance of chronic pain: Component analysis and a revised assessment method. *Pain*, *107*(1), 159–166. https://doi.org/10.1016/j.pain.2003.10.012

McLoughlin, M. J., Colbert, L. H., Stegner, A. J., & Cook, D. B. (2011). Are women with fibromyalgia less physically active than healthy women? *Medicine & Science in Sports and Exercise*, *43*(5), 905–912. https://doi.org/10.1249/MSS.0b013e3181fca1ea

McWilliams, L. A., Cox, B. J., & Enns, M. W. (2003). Mood and anxiety disorders associated with chronic pain: An examination in a nationally representative sample. *Pain*, *106*(1), 127–133. https://doi.org/10.1016/S0304-3959(03)00301-4

MedCalc Statistical Software. (2020). Ostend, Belgium: MedCalc Software Ltd. Retrieved from https://www.medcalc.org

Meeus, M., Nijs, J., Van Wilgen, P., Noten, S., Goubert, D., & Huijnen, I. (2016). Moving on to movement in patients with chronic joint pain. In *Pain: Clinical Updates* (Vol. 1). Retrieved from https://painsa.org.za/wp-content/uploads/2016/11/Moving-onto-Movement-in-Pts-w-Chronic-Jt-Pain-IASP.pdf

Merskey, H., & Bogduk, N. (1994). *Classification of chronic pain: Descriptions of chronic pain syndromes and definitions of pain terms* (2nd ed). Seattle, WA, USA: IASP Press.

Meulders, A., Harvie, D. S., Bowering, J. K., Caragianis, S., Vlaeyen, J. W. S., & Moseley, G. L. (2014). Contingency learning deficits and generalization in chronic unilateral hand pain patients. *The Journal of Pain*, *15*(10), 1046–1056. https://doi.org/10.1016/j.jpain.2014.07.005

Miles, S., Howlett, C. A., Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, *53*, 2083–2091. https://doi.org/10.3758/s13428-021-01551-3

Miles, S., Nedeljkovic, M., Sumner, P., & Phillipou, A. (2022). Understanding self-report and neurocognitive assessments of cognitive flexibility in people with and without lifetime anorexia nervosa. *Cognitive Neuropsychiatry*, *27*(5), 325–341. https://doi.org/10.1080/13546805.2022.2038554

Milligan, E. D., & Watkins, L. R. (2009). Pathological and protective roles of glia in chronic pain. *Nature Reviews Neuroscience*, *10*, 23–36. https://doi.org/10.1038/nrn2533

Millisecond Software. (2021). *Inquisit 6*. Retrieved from http://www.millisecond.com

Mills, S. E. E., Nicolson, K. P., & Smith, B. H. (2019). Chronic pain: A review of its epidemiology and associated factors in population-based studies. *British Journal of Anaesthesia*, *123*(2), e273–e283. https://doi.org/10.1016/j.bja.2019.03.023

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology*, *9*(1), 90–100. https://doi.org/10.1001/archneur.1963.00460070100010

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, *21*(2), 169–183. https://doi.org/10.1055/s-2000-7563

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mogil, J. S. (2012). Sex differences in pain inhibition: multiple explanations of a controversial phenonmenon. *Nature Reviews Neuroscience*, *13*, 859–866. https://doi.org/10.1038/nrn3360

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), p.e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moriarty, O., Ruane, N., O'Gorman, D., Maharaj, C. H., Mitchell, C., Sarma, K. M., … McGuire, B. E. (2017). Cognitive impairment in patients with chronic neuropathic or radicular pain: An interaction of pain and age. *Frontiers in Behavioral Neuroscience*, *11*(100). https://doi.org/10.3389/fnbeh.2017.00100

Moseley, G. L. (2003). A pain neuromatrix approach to patients with chronic pain. *Manual Therapy, 8*(3), 130-140. https://doi.org/10.1016/S1356-689X(03)00051-1

Moseley, G. L., & Arntz, A. (2007). The context of a noxious stimulus affects the pain it evokes. *Pain*, *133*(1), 64–71. https://doi.org/10.1016/j.pain.2007.03.002

Moseley, G. L., & Butler, D. S. (2015). Fifteen years of explaining pain: The past, present and future. *The Journal of Pain*, *16*(9), 807–813. https://doi.org/10.1016/j.jpain.2015.05.005

Moseley, G. L., & Butler, D. S. (2017). *Explain pain supercharged*. Adelaide: Noigroup Publications.

Moseley, G. L., & Vlaeyen, J. S. (2015). Beyond nociception: The imprecision hypothesis of chronic pain. *Pain*, *156*(1), 35–38. https://doi.org/10.1016/j.pain.0000000000000014

Moseley, L. (2002). Combined physiotherapy and education is efficacious for chronic low back pain. *Australian Journal of Physiotherapy*, *48*(4), 297–302. https://doi.org/10.1016/S0004-9514(14)60169-0

Moseley, L. (2003). Unraveling the barriers to reconceptualization of the problem in chronic pain: The actual and perceived ability of patients and health professionals to understand the neurophysiology. *The Journal of Pain*, *4*(4), 184–189. https://doi.org/10.1016/S1526-5900(03)00488-7

Müller, U., & Kerns, K. (2015). The development of executive function. In *In L. S. Liben, U. Müller, & R. M. Lerner (Eds.), Handbook of child psychology and developmental science: Cognitive processes* (pp. 571–623). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy214

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(0021). https://doi.org/10.1038/s41562-016-0021

Nadar, M. S., Jasem, Z., & Manee, F. S. (2016). The cognitive functions in adults with chronic pain: A comparative study. *Pain and Research Management*, *2016*(5719380), 1–8. https://doi.org/10.1155/2016/5719380

Nahin, R. L. (2015). Estimates of pain prevalence and severity in adults: United States, 2012. *The Journal of Pain*, *16*(8), 769–780. https://doi.org/10.1016/j.jpain.2015.05.002

NHLI. (2014). Quality assessment tool for observational cohort and cross-sectional studies. Retrieved from https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Okuda, T., Asano, K., Numata, N., Hirano, Y., Yamamoto, T., Tanaka, M., … Nakazato, M. (2017). Feasibility of cognitive remediation therapy for adults with autism spectrum disorders: A single-group pilot study. *Neuropsychiatric Disease and Treatment*, *13*, 2185–2191. https://doi.org/10.2147/NDT.S141555

Ongaro, G., & Kaptchuk, T. J. (2019). Symptom perception, placebo effects, and the Bayesian brain. *Pain*, *160*(1), 1–4. https://doi.org/10.1097/j.pain.0000000000001367

Page, M. J., McKenzie, J. E., & Higgins, J. P. T. (2018). Tools for assessing risk of reporting biases in studies and syntheses of studies: A systematic review. *BMJ Open*, *8*(3), e019703. https://doi.org/10.1136/bmjopen-2017-019703

Palm, K. M., & Follette, V. M. (2011). The roles of cognitive flexibility and experiential avoidance in explaining psychological distress in survivors of interpersonal victimization. *Journal of Psychopathology and Behavioral Assessment*, *33*, 79–86. https://doi.org/10.1007/s10862-010-9201-x

Pang, E. W., Dunkley, B. T., Doesburg, S. M., da Costa, L., & Taylor, M. J. (2016). Reduced brain connectivity and mental flexibility in mild traumatic brain injury. *Annals of Clinical and Translational Neurology*, *3*(2), 124–131. https://doi.org/10.1002/acn3.280

Parr, T., Rees, G., & Friston, K. J. (2018). Computational neuropsychology and bayesian inference. *Frontiers in Human Neuroscience*, *12*(61). https://doi.org/10.3389/fnhum.2018.00061

Pavlov, I. P. (2010). Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences*, *17*(3), 136–141. https://doi.org/10.5214/ans.0972-

7531.1017309

Pitcher, M. H., Von Korff, M., Bushnell, M. C., & Porter, L. (2019). Prevalence and profile of high-impact chronic pain in the United States. *The Journal of Pain*, *20*(2), 146–160. https://doi.org/10.1016/j.jpain.2018.07.006

Ploner, M., Sorg, C., & Gross, J. (2017). Brain rhythms of pain. *Trends in Cognitive Sciences*, *21*(2), 100–110. https://doi.org/10.1016/j.tics.2016.12.001

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65. https://doi.org/10.1016/j.acn.2004.02.005

Rabin, L. A., Roth, R. M., Isquith, P. K., Wishart, H. A., Nutter-Upham, K. E., Pare, N., … Saykin, A. J. (2006). Self- and informant reports of executive function on the BRIEF-A in MCI and older adults with cognitive complaints. *Archives of Clinical Neuropsychology*, *21*(7), 721–732. https://doi.org/10.1016/j.acn.2006.08.004

Raja, S. N., Carr, D. B., Cohen, M., Finnerup, N. B., H, F., Gibson, S., … Vader, K. (2020). The revised International Association for the Study of Pain definition of pain: Concepts, challenges and compromises. *Pain*, *161*(9), 1976–1982. https://doi.org/10.1097/j.pain.0000000000001939

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Rathbone, M., Parkinson, W., Rehman, Y., Jiang, S., Bhandari, M., & Kumbhare, D. (2016). Magnitude and variability of effect sizes for the associations between chronic pain and cognitive test performances: A meta-analysis. *British Journal of Pain*, *10*(3), 141–155. https://doi.org/10.1177/2049463716642600

Reimer, C. B., Chen, Z., Bundt, C., Eben, C., London, R. E., & Vardanian, S. (2019). Open up – the mission statement of the Control of Impulsive Action (Ctrl-ImpAct) Lab on open science. *Psychologica Belgica*, *59*(1), 321–337. https://doi.org/10.5334/pb.494

Reitan, R. M. (1955). The relation of the Trail Making Test to organic brain damage. *Journal of Consulting Psychology*, *19*(5), 393–394. https://doi.org/10.1037/h0044509

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*(3), 271–276. https://doi.org/10.2466/pms.1958.8.3.271

Reneman, M. F., Kleen, M., Trompetter, H. R., Schiphorst Preuper, H. R., Köke, A., van Baalen, B., & Schreurs, K. M. G. (2014). Measuring avoidance of pain: Validation of the Acceptance and Action Questionnaire II-pain version. *International Journal of Rehabilitation Research*, *37*(2), 125–129. https://doi.org/10.1097/MRR.0000000000000044

Reynolds, C. R. (2002). *Comprehensive Trail-Making Test (CTMT): Examiner's manual*. Austin, TX: PRO-ED.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., Lawrence, A. D., Mcinnes, L., & Rabbitt, P. M. A. (1998). A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. *Journal of the International Neuropsychological Society*, *4*(5), 474–490. https://doi.org/10.1017/S1355617798455073

Roberts, M. E., Barthel, F. M. S., Lopez, C., Tchanturia, K., & Treasure, J. L. (2011). Development and validation of the Detail and Flexibility Questionnaire (DFlex) in eating disorders. *Eating Behaviors*, *12*(3), 168–174. https://doi.org/10.1016/j.eatbeh.2011.04.001

Roth, R. M., Isquith, P. K., & Gioia, G. A. (2005). *Behavior Rating Inventory of Executive Function -- Adult Version*. Lutz, Florida: Psychological Assessment Resources, Inc.

Sá, K. N., Moreira, L., Baptista, A. F., Yeng, L. T., Teixeira, M. J., Galhardoni, R., & de Andrade, D. C. (2019). Prevalence of chronic pain in developing countries: Systematic review and meta-analysis. *Pain Reports*, *4*(6), e779. https://doi.org/10.1097/PR9.0000000000000779

Şahin, M. D., & Aybek, E. C. (2019). Jamovi: An easy to use statistical software for the social scientists. *International Journal of Assessment Tools in Education*, *6*(4), 670–692. https://doi.org/10.21449/ijate.661803

Sanderson, S., Tatt, I. D., & Higgins, J. P. T. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, *36*(3), 666–676. https://doi.org/10.1093/ije/dym018

Schofield, D. J., Shrestha, R. N., Passey, M. E., Earnest, A., & Fletcher, S. L. (2008). Chronic disease and labour force participation among older Australians. *The Medical Journal of Australia*,

189(8), 447–450. https://doi.org/10.5694/j.1326-5377.2008.tb02119.x

Schofield, D. J., Shrestha, R. N., Percival, R., Callander, E. J., Kelly, S. J., & Passey, M. E. (2011). Early retirement and the financial assests of individuals with back problems. *European Spine Journal*, *20*, 731–736. https://doi.org/10.1007/s00586-010-1647-8

Schultz, P. W., & Searleman, A. (2002). Rigidity of thought and behavior: 100 years of research. *Genetic, Social, and General Psychology Monographs*, *128*(2), 165–207.

Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry*, *25*(4), 405–414. https://doi.org/10.2307/2785779

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, *349*, g7647. https://doi.org/10.1136/bmj.g7647

Soto, E. F., Kofler, M. J., Singh, L. J., Wells, E. L., Irwin, L. N., Groves, N. B., & Miller, C. E. (2020). Executive functioning rating scales: Ecologically valid or construct invalid? *Neuropsychology*, *34*(6), 605–619. https://doi.org/10.1037/neu0000681

Spencer, R. J., Drag, L. L., Walker, S. J., & Bieliauskas, L. A. (2010). Self-reported cognitive symptoms following mild traumatic brain injury are poorly associated with neuropsychological performance in OIF/OEF veterans. *Journal of Rehabilitation Research & Development*, *47*(6), 521–530. https://doi.org/10.1682/JRRD.2009.11.0181

St. Sauver, J. L., Warner, D. O., Yawn, B. P., Jacobson, D. J., McGree, M. E., Pankratz, J. J., … Rocca, W. A. (2013). Why patients visit their doctors: Assessing the most prevalent conditions in a defined American population. *Mayo Clinic Proceedings*, *88*(1), 56–67. https://doi.org/10.1016/j.mayocp.2012.08.020

Stedal, K., & Dahlgren, C. L. (2015). Neuropsychological assessment in adolescents with anorexia nervosa – exploring the relationship between self-report and performance-based testing. *Journal of Eating Disorders*, *3*(27), 1–12. https://doi.org/10.1186/s40337-015-0062-2

Steffens, D., Hancock, M. J., Maher, C. G., Williams, C., Jensen, T. S., & Latimer, J. (2014). Does magnetic resonance imaging predict future low back pain? A systematic review. *European Journal of Pain*, *18*(6), 755–765. https://doi.org/10.1002/j.1532-2149.2013.00427.x

Steinglass, J. E., Walsh, B. T., & Stern, Y. (2006). Set shifting deficit in anorexia nervosa. *Journal of the International Neuropsychological Society*, *12*(3), 431–435.

https://doi.org/10.1017/S1355617706060528

Steinmetz, J.-P., Brunner, M., Loarer, E., & Houssemand, C. (2010). Incomplete psychometric equivalence of scores obtained on the manual and the computer version of the Wisconsin Card Sorting Test? *Psychological Assessment*, *22*(1), 199–202. https://doi.org/10.1037/a0017661

Stewin, L. (1983). The concept of rigidity: An enigma. *International Journal for the Advancement of Counselling*, *6*, 227–232. https://doi.org/10.1007/BF00124273

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Suhr, J. A. (2003). Neuropsychological impairment in fibromyalgia: Relation to depression, fatigue, and pain. *Journal of Psychosomatic Research*, *55*(4), 321–329. https://doi.org/10.1016/S0022-3999(02)00628-1

Suurmond, R., van Rhee, H., & Hak, T. (2017). Introduction, comparison, and validation of Meta-Essentials: A free and simple tool for meta-analysis. *Research Synthesis Methods*, *8*, 537–553. https://doi.org/10.1002/jrsm.1260

Tabor, A., & Burr, C. (2019). Bayesian learning models of pain: A call to action. *Current Opinion in Behavioral Sciences*, *26*, 54–61. https://doi.org/10.1016/j.cobeha.2018.10.006

Tabor, A., Thacker, M. A., Moseley, G. L., & Kording, K. P. (2017). Pain: A statistical account. *PLOS Computational Biology*, *13*(1), e1005142. https://doi.org/10.1371/journal.pcbi.1005142

Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., … Morris, R. (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PLoS ONE*, *7*(1), e28331. https://doi.org/10.1371/journal.pone.0028331

Tchanturia, K., Lloyd, S., & Lang, K. (2013). Cognitive remediation therapy for anorexia nervosa: Current evidence and future research directions. *International Journal of Eating Disorders*, *46*(5), 492–495. https://doi.org/10.1002/eat.22106

Tchanturia, K., Morris, R. G., Anderluh, M. B., Collier, D. A., Nikolaou, V., & Treasure, J. (2004). Set shifting in anorexia nervosa: An examination before and after weight gain, in full recovery and relationship to childhood and adult OCPD traits. *Journal of Psychiatric Research*, *38*(5), 545–552. https://doi.org/10.1016/j.jpsychires.2004.03.001

Tchanturia, K., Morris, R. G., Surguladze, S., & Treasure, J. (2002). An examination of perceptual and cognitive set shifting tasks in acute anorexia nervosa and following recovery. *Eating and Weight Disorders*, *7*, 312–315. https://doi.org/10.1007/BF03324978

Tchanturia, K., Serpell, L., Troop, N., & Treasure, J. (2001). Perceptual illusions in eating disorders: Rigid and fluctuating styles. *Journal of Behavior Therapy and Experimental Psychiatry*, *32*(3), 107–115. https://doi.org/10.1016/S0005-7916(01)00025-8

Tesio, V., Torta, D. M. E., Colonna, F., Leombruni, P., Ghiggia, A., Fusaro, E., … Castelli, L. (2015). Are fibromyalgia patients cognitively impaired? Objective and subjective neuropsychological evidence. *Arthritis Care & Research*, *67*(1), 143–150. https://doi.org/10.1002/acr.22403

The Cochrane Collaboration, (2020). *Review Manager (RevMan) [Computer program]*.

Thoma, P., Wiebel, B., & Daum, I. (2007). Response inhibition and cognitive flexibility in schizophrenia with and without comorbid substance use disorder. *Schizophrenia Research*, *92*(1–3), 168–180. https://doi.org/10.1016/j.schres.2007.02.004

Thorndike, E. L. (1913). *An introduction to the theory of mental and social measurements* (2nd ed). New York: Teachers College Press.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Traxler, J., Madden, V. J., Moseley, G. L., & Vlaeyen, J. W. S. (2019). Modulating pain thresholds through classical conditioning. *PeerJ*, *7*, e6486. https://doi.org/10.7717/peerj.6486

Treede, R.-D., Rief, W., Barke, A., Aziz, Q., Bennett M I, Benoliel, R., … Wang, S.-J. (2015). A classification of chronic pain for ICD-11. *Pain*, *156*(6), 1003–1007. https://doi.org/10.1097/j.pain.0000000000000160

Turk, D. C., Fillingim, R. B., Ohrbach, R., & Patel, K. V. (2016). Assessment of psychosocial and functional impact of chronic pain. *The Journal of Pain*, *17*(9), T21–T49. https://doi.org/10.1016/j.jpain.2016.02.006

Tyndall, I., Waldeck D, Pancani, L., Whelan, R., Roche, B., & Dawson, D. L. (2019). The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *Journal of Contextual Behavioral Science*, *12*, 278–284. https://doi.org/10.1016/j.jcbs.2018.09.005

Uddin, L. Q. (2021). Cognitive and behavioural flexibility: Neural mechanisms and clinical considerations. *Nature Reviews Neuroscience*, *22*, 167–179. https://doi.org/10.1038/s41583-021-00428-w

Uznadze, D. N. (1966). *The psychology of set*. New York: Consultants Bureau.

van Hecke, O., Torrance, N., & Smith, B. H. (2013). Chronic pain epidemiology - where do lifestyle factors fit in? *British Journal of Pain*, *7*(4), 209–217. https://doi.org/10.1177/2049463713493264

van Leeuwen, M. T., Blyth, F. M., March, L. M., Nicholas, M. K., & Cousins, M. J. (2006). Chronic pain and reduced work effectiveness: The hidden cost to Australian employers. *European Journal of Pain*, *10*(2), 161–166. https://doi.org/10.1016/j.ejpain.2005.02.007

Vance, J. (2017). Action prevents error - predictive processing without active inference. In *In T. Metzinger & W. Wiese (Eds.). Philosophy and Predictive Processing: 25.* (pp. 1–14). Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958573260

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., … Egger, M. (2014). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *International Journal of Surgery*, *12*, 1500–1524. https://doi.org/10.1016/j.ijsu.2014.07.014

Verdejo-Garcia, A., Lopez-Torrecillas, F., Calandre, E. P., Delgado-Rodriguez, A., & Bechara, A. (2009). Executive function and decision-making in women with fibromyalgia. *Archives of Clinical Neuropsychology*, *24*(1), 113–122. https://doi.org/10.1093/arclin/acp014

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vila, M. R., Todorovic, M. S., Tang, C., Fisher, M., Steinberg, A., Field, B., … Haroutounian, S. (2020). Cognitive flexibility and persistent post-surgical pain: The FLEXCAPP prospective observational study. *British Journal of Anaesthesia*, *124*(5), 614–622. https://doi.org/10.1016/j.bja.2020.02.002

Vlaeyen, J. W. S. (2014). Psychological flexibility: What theory and which predictions? *The Journal of Pain*, *15*(3), 235–236. https://doi.org/10.1016/j.jpain.2014.01.003

Walker, B. F. (1999). The prevalence of low back pain: A systematic review of the literature from 1966 to 1998. *Asia-Pacific Journal of Public Health*, *11*(1), 45–51.

https://doi.org/10.1177/101053959901100110

Walker, B. F., Muller, R., & Grant, W. D. (2003). Low back pain in Australian adults: The economic burden. *Asia-Pacific Journal of Public Health*, *15*(2), 79–87. https://doi.org/10.1177/101053950301500202

Wall, P. D., & McMahon, S. B. (1986). The relationship of perceived pain to afferent nerve impulses. *Trends in Neurosciences*, *9*, 254–255. https://doi.org/10.1016/0166-2236(86)90070-6

Wand, B. M., Parkitny, L., O'Connell, N. E., Luomajoki, H., McAuley, J. H., Thacker, M., & Moseley, G. L. (2011). Cortical changes in chronic low back pain: Current state of the art and implications for clinical practice. *Manual Therapy*, *16*(1), 15–20. https://doi.org/10.1016/j.math.2010.06.008

Wang, T., Li, M., Xu, S., Liu, B., Wu, T., Lu, F., … Wang, J. (2019). Relations between trait anxiety and depression: A mediated moderation model. *Journal of Affective Disorders*, *244*, 217–222. https://doi.org/10.1016/j.jad.2018.09.074

Wecker, N. S., Kramer, J. H., Hallam, B. J., & Delis, D. C. (2005). Mental flexibility: Age effects on switching. *Neuropsychology*, *19*(3), 345–352. https://doi.org/10.1037/0894-4105.19.3.345

Weiner, D. K., Rudy, T. E., Morrow, L., Slaboda, J., & Lieber, S. (2006). The relationship between pain, neuropsychological performance, and physical function in community-dwelling older adults with chronic low back pain. *Pain Medicine*, *7*(1), 60–70. https://doi.org/10.1111/j.1526-4637.2006.00091.x

Wennerhold, L., & Friese, M. (2020). Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra: Psychology*, *6*(1), 9. https://doi.org/10.1525/collabra.276

Westwood, H., Stahl, D., Mandy, W., & Tchanturia, K. (2016). The set-shifting profiles of anorexia nervosa and autism spectrum disorder using the Wisconsin Card Sorting Test: A systematic review and meta-analysis. *Psychological Medicine*, *46*(9), 1809–1827. https://doi.org/10.1017/S0033291716000581

Wicksell, R. K., Renöfält, J., Olsson, G. L., Bond, F. W., & Melin, L. (2008). Avoidance and cognitive fusion - Central components in pain related disability? Development and preliminary validation of the Psychological Inflexibility in Pain Scale (PIPS). *European Journal of Pain*, *12*(4), 491–500. https://doi.org/10.1016/j.ejpain.2007.08.003

Wiebe, D. J., Baker, A. C., Suchy, Y., Stump, T. K., & Berg, C. A. (2018). Individual differences and day-to-day fluctuations in goal planning and type 1 diabetes management. *Health Psychology*, *37*(7), 638–646. https://doi.org/10.1037/hea0000624

Wiech, K. (2016). Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science*, *354*(6312), 584–587. https://doi.org/10.1126/science.aaf8934

Williams, A. C. D. C., & Craig, K. D. (2016). Updating the definition of pain. *Pain*, *157*(11), 2420–2423. https://doi.org/10.1097/j.pain.0000000000000613

Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behavior Therapy*, *45*(6), 831–839. https://doi.org/10.1016/j.beth.2014.07.002

Wollenhaupt, C., Wilke, L., Erim, Y., Rauh, M., Steins-Loeber, S., & Paslakis, G. (2019). The association of leptin secretion with cognitive performance in patients with eating disorders. *Psychiatry Research*, *276*, 269–277. https://doi.org/10.1016/j.psychres.2019.05.001

Yeniad, N., Malda, M., Mesman, J., van Ijzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences*, *23*, 1–9. https://doi.org/10.1016/j.lindif.2012.10.004

Yu, Y., Yu, Y., & Lin, Y. (2020). Anxiety and depression aggravate impulsiveness: The mediating and moderating role of cognitive flexibility. *Psychology, Health & Medicine*, *25*(1), 25–36. https://doi.org/10.1080/13548506.2019.1601748

Zeng, X., Zhang, Y., Kwong, J. S. W., Zhang, C., Li, S., Sun, F., … Du, L. (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: A systematic review. *Journal of Evidence-Based Medicine*, *8*(1), 2–10. https://doi.org/10.1111/jebm.12141

Zetterqvist, V., Holmström, L., Maathz, P., & Wicksell, R. K. (2017). Pain avoidance predicts disability and depressive symptoms three years later in individuals with whiplash complaints. *Acta Anaesthesiologica Scandinavica*, *61*(4), 445–455. https://doi.org/10.1111/aas.12874

Zheng, Z. (2021). Reasons, challenges, and some tools for doing reproducible transportation research. *Communications in Transportation Research*, *1*, 100004. https://doi.org/10.1016/j.commtr.2021.100004

CHAPTER EIGHT

Research dissemination: Publications, presentations, awards, impact and additional activities

8.1 Research outputs

First author publications arising from this thesis

Under review

1. **Howlett, C. A.,** Stanford, T., Berryman, C., Karran, E. L., Bellan, V., Coussens, S., Miles, S., & Moseley, G.L. Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study.

Published in peer-reviewed journals

*Co-first authors – both authors contributed equally to the publication.
#Co-last authors – both authors contributed equally to the publication.

1. ***Howlett, C. A.,** *Miles, S., Berryman, C., #Phillipou, A., & #Moseley, G.L. (2023). Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy. *Australian Journal of Psychology, 75*(1), 2174684. https://doi.org/10.1080/00049530.2023.2174684

2. **Howlett, C. A.,** Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., Karran, E. L., Szeto, K., Pronk, L., Miles, S., & Moseley, G. L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology, 36*(5), 347-372. https://doi.org/10.1037/neu0000796

3. **Howlett, C. A.,** Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., Karran, E. L., Szeto, K., Pronk, L., Miles, S., & Moseley, G. L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review, 88*(102061), 1-21. https://doi.org/10.1016/j.cpr.2021.102061

4. *Miles, S., ***Howlett, C. A.,** Berryman, C., Nedeljkovic, M., #Moseley, G. L., & #Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, 53, 2083-2091. https://doi.org/10.3758/s13428-021-01551-3

Co-author publications produced during candidature, but not included in this thesis

Published in peer-reviewed journals

1. Liebherr, M., Corcoran, A. W., Alday, P. M., Coussens, S., Bellan, V., **Howlett, C. A.,** Immink, M. A., Kohler, M., Schlesewsky, M., & Bornkessel-Schlesewsky, I. EEG and behavioural correlates of attentional processing while walking and navigating naturalistic environments. *Scientific Reports, 11*(1), 1-13. https://doi.org/10.1101/2021.05.27.445993

2. Bornkessel-Schlesewsky, I., Sharrad, I., **Howlett, C. A.,** Alday, P.M., Corcoran, A. W., Bellan, V., Wilkinson, E., Kliegl, R., Lewis, R. L., Small, S.L., & Schlesewsky, M. (2022). Rapid adaptation of predictive models during language comprehension: Aperiodic EEG slope, individual alpha frequency and idea density modulate individual differences in real-time model updating. *Frontiers in Psychology, 13*(817516), https://doi.org/10.3389/fpsyg.2022.817516

8.2 Presentations during candidature

Oral abstract presentations (peer reviewed):

2022 **Australian Pain Society 42nd Annual Scientific Meeting** (April 10-13)
Topical Session: *'Goal setting, cognitive flexibility, and their relationship to adjustment in chronic pain'*
Co-speakers: Associate Professor Toby Newton-John (Chair), Tim Austin and Lawrence Roux
Hotel Grand Chancellor, Hobart, Tasmania, Australia

Poster abstract presentations (peer reviewed):

2023 **International Convention of Psychological Science** (March 09-11)
Presentation title: *'Evaluating the relationship between self-report and neuropsychological assessments of cognitive flexibility in clinical and non-clinical cohorts'*
Co-authors: MA Wewege, C Berryman, A Oldach, E Jennings, E Moore, EL Karran, K Szeto, L Pronk, S Miles, GL Moseley
The SQUARE – Brussels Convention Centre, Brussels, Belgium

***2020** **International Association for the Study of Pain (IASP) World Congress on Pain** (August 04-08)
Presentation title: '*Evaluating the impact of video and text-based online delivery of pain education'*
Co-authors: EL Karran, TE Ding, RJ Bray, GL Moseley, MJ Catley
Exhibit Hall, RAI Amsterdam, Amsterdam, The Netherlands

*Withdrawn because of the COVID-19 pandemic

Other oral presentations:

2021 **Pain Adelaide Conference** (October 18)
Session: PhD presentations
Presentation title: *'Untangling the construct of cognitive flexibility'* (5 minutes)
National Wine Centre of Australia, Adelaide, South Australia, Australia

8.3 Scholarships and awards attained during candidature

***2020** **Financial Aid Award**
Free registration to the 2020 IASP World Congress on Pain and $960 USD towards Travel and Accommodation Expenses, International Association for the Study of Pain (IASP) World Congress on Pain, Amsterdam, The Netherlands

2020 **2nd Prize, Research Week Poster Competition**, Allied Health and Human Performance, University of South Australia, Adelaide, South Australia, Australia

2018 **Australian Research Training Program domestic (RTPd) Stipend (~$81, 246 AUD) and Tuition Offset**

*Unused because of the COVID-19 pandemic

8.4 Impact of studies contained in this thesis

The work contained in this thesis has led to four published journal articles, totalling 89 citations, and one manuscript that is currently under review at a peer-reviewed journal.

The work presented in Chapter 2 and published in *Behavior Research Methods* has a total of 58 citations and an Altmetric score of 21, situating it in the 89th percentile (ranked 43, 462nd) of the >400,000 tracked articles of a similar age in all journals and the 86th percentile (ranked 8th) of the 50 tracked articles of a similar age. This work was the result of a collaboration with colleagues from Swinburne University, which aimed to provide clarification on various aspects of the WCST. It led to wholesale changes of the WCST on one of the most widely used online software programs for cognitive, social and psychological experiments and direct consultation with the software designers. The software is used by over 3000 leading research institutions worldwide. As such, the work contained in Chapter 2 has directly altered the investigation of cognitive flexibility at over 3000 research institutions worldwide. The work presented in Chapter 3 and published in *Clinical Psychology Review* has a total of 23 citations; the work presented in Chapter 4 and published in *Neuropsychology* has a total of seven citations; the work presented in Chapter 5 and published in the *Australian Journal of Psychology* has one citation.

The work contained in this thesis also resulted in an invitation to present a topical workshop at the *Australian Pain Society 42nd Annual Scientific Meeting* (2022). The studies presented in Chapters 3 and 4 were presented as a scientific poster at the *International Convention of Psychological Science* conference (2023).

8.5 Additional activities during candidature

Manuscript reviews for peer review journal(s):

2022 Advances in Cognitive Psychology (1)

Community outreach and engagement

2021 Pain Revolution "Go the Distance Challenge" (October 1-31): walked/ran 31kms to raise funds to support people in pain (and their clinical support teams) in rural and remote communities in Australia

2019 Volunteer/Organiser – The Colour the World Orange CRPS Awareness Night, University of South Australia, Adelaide, South Australia, Australia

2019 Blog post – Noijam (Noigroup): 'Predictive Processing: A potential theory for persistent pain, and the power of discrepancy in facilitating change' (at the time of submission, the blog post had a total of 8,459 views in over 25 countries and had been viewed in every continent except Antarctica, placing it in the top 20 of all-time views on Noijam)
https://www.noigroup.com/noijam/predictive-processing-a-potential-theory-for-persistent-pain-and-the-power-of-discrepancy-in-facilitating-change/

2018 Science Correspondent for Neuro Orthopaedic Institute (NOI) Explain Pain 3-day workshop (November 09-11)
Melbourne Cricket Ground (MCG), Melbourne, Victoria, Australia

2018 Volunteer/Organiser – The Colour the World Orange CRPS Awareness Night, University of South Australia, Adelaide, South Australia, Australia

Employment

Online Casual Tutor, University of South Australia

2022 Introduction to Evidence-Based Practice and Research in Health Sciences

2020 Psychology 1B

Research assistant, University of South Australia

2019 Cognitive Neuroscience Laboratory – Australian Research Centre for Interactive and Virtual Environments

CHAPTER NINE

Appendices

List of Appendices

- Appendix A: Supplementary material for Chapters Three and Four
- Appendix B: Supplementary material for Chapter Three
- Appendix C: Supplementary material for Chapter Four
- Appendix D: Supplementary material for Chapter Six
- Appendix E: Permissions
- Appendix F: Statements of Authorship

Appendix A: Supplementary material for Chapters Three and Four

1. **Howlett, C.A.,** Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, 88(102061). https://doi.org/10.1016/j.cpr.2021.102061

2. **Howlett, C.A.,** Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology*, *36*(5), 347-372. https://doi.org/10.1037/neu0000796

Outline:

- **Supplementary 1 (S1):** Pre-registered protocol for Chapters Three and Four
- **Supplementary 2 (S2):** Search strategy for Chapters Three and Four

Supplementary 1 (S1): Pre-registered protocol Chapters Three and Four

Registered on Open Science Framework on 19th November 2019: www.doi.org/10.17605/OSF.IO/6QS2U and on PROSPERO (registration number: CRD42019131567)

1. **Title:** Evaluating the relationship between self-report tests and neuropsychological tests assessing flexibility: a protocol for a systematic review

2. **Investigator details:**

1. Caitlin A. Howlett, PhD Candidate (BA [Psych]), School of Health Sciences, University of South Australia; caitlin.howlett@mymail.unisa.edu.au; 08 8302 1524
2. Prof Ina Bornkessel-Schlesewsky, Prof in Cognitive Neuroscience (MA [Arts in Linguistics], PhD), School of Psychology, Social Work and Social Policy; University of South Australia; ina.bornkessel-schlesewsky@unisa.edu.au; 08 83024236
3. Dr Valeria Bellan, Research Fellow (BA [Psych], MA [Clin Psych], PhD), School of Psychology, Social Work and Social Policy; University of South Australia; valeria.bellan@unisa.edu.au; 08 83024298
4. Dr Emma Karran, Research Fellow (BAppSc Phty [Hons], Grad Dip [Psych], PhD), School of Health Sciences, University of South Australia; emma.karran@unisa.edu.au; 08 8302 1128
5. Dr Carolyn Berryman, NHMRC Early Career Fellow and Adjunct Research Fellow, (BAppSc [Phty], MAppSc Phty [Hons], MSc, Med [Pain Mgt], PhD), School of Health Sciences, University of South Australia; carolyn.berryman@unisa.edu.au; 08 8302 1290
6. Prof G. Lorimer Moseley, Prof in Clinical Neurosciences and Chair in Physiotherapy (BAppSc [Phty], PhD, DSc, FACP, FAAHMS), School of Health Sciences, University of South Australia; lorimer.moseley@unisa.edu.au; 08 8302 1416

3. Background

Persistent pain is a global health concern[1], characterised by pain that continues beyond the typical timeframe required for healing after sustaining tissue damage (i.e., lasting more than 3 months)[2]. Over 3 million Australians are thought to live with persistent pain, with the prevalence expected to increase to 5 million people by 2050 as a result of the nation's aging population[3]. The economic burden of persistent pain is substantial, with estimates suggesting annual costs of over $34 billion in Australia[3] and between $560 billion and $635 billion in the United States – exceeding the combined cost of cardiovascular disease and cancer[4].

Contemporary understanding of pain emphasises that it is complex and multifactorial. However, despite considerable progress, much remains to be understood - particularly what factors determine whether someone develops persistent pain and what may well prevent the transition from acute to persistent pain. One possibility lies in potentially important links between persistent pain and cognitive flexibility, one element of higher-order cognitive abilities, or executive function[5]. Cognitive flexibility is thought to be an emergent property of the efficient cognitive system or the interaction between several cognitive processes including attention, working memory, inhibition and switching[6,7]. Thus, it can be broadly described as the ability to alter one's

perspective, thinking styles, and cognitive and behavioural strategies in order to adapt to changing demands[5].

However, the literature in the field seems to be characterised by both confidence in the discrete and assessable nature of cognitive flexibility, and a lack of a unified understanding of what flexibility is as a construct. One review paper[7] identified that cognitive flexibility is poorly conceptualised in the literature. To illustrate: some authors define flexibility as a specific cognitive ability or skill (i.e. predominately investigated with "set shifting" or "task-switching" paradigms)[6]; others define it as a cognitive state (i.e. cognitive rigidity or the inability to re-appraise maladaptive cognitions)[8]. Additionally, Ionescu (2012)[7] grouped various studies that defined cognitive flexibility as: (1) a higher-order ability, such as cognitive control; (2) a property of mental states; and (3) a measure of divergent thinking, demonstrating the lack of consensus and broad spectrum with which cognitive flexibility has been defined. The literature in the field of cognitive flexibility has had a tendency to (1) use the terms "cognitive", "mental" or "psychological" flexibility synonymously and (2) assume that self-report and neuropsychological tests assess the same underlying aspects of flexibility and can be used interchangeably. In fact, there is evidence to suggest that there is poor correspondence between these two forms of tests[9,10]. This review will avoid confusing terminology and adopt the term 'flexibility' throughout, and clearly delineate between assessments using self-report and neuropsychological tests because investigating whether there is a relationship between the two will deepen our understanding of and ability to interpret critical aspects of flexibility.

The most popular self-report instruments used to assess flexibility include the Cognitive Flexibility Scale (CFS)[11], the Cognitive Flexibility Inventory (CFI)[8] and the Attributional Style Questionnaire[12]. Alternatively, the most widely used neuropsychological instruments used to assess flexibility include the Stroop Colour and Word Test[13], the Trail Making Test (TMT)[14], the Wisconsin Card Sorting Test (WCST)[15] and the Alternate Uses Task[16]. Currently, there is little consensus regarding the optimal way to assess flexibility, and thus the instruments used to measure it are largely diverse and how well correlated these tests are remains unknown.

It is clear that flexibility is an important higher-order property of executive function that allows us to complete complex tasks. There is sufficient evidence to suggest that lower levels of flexibility are prevalent in various populations including depression[17,18], anxiety[19], eating disoders[20,21] and persistent pain[22,23], highlighting the important clinical implications (i.e. diagnostic and for treatment) in understanding the relationship between self-report and neuropsychological tests used to assess flexibility. However, to date, there is no published systematic analysis of all the available literature. Additionally, it remains unknown whether the body of literature has administered these two forms of assessments together at a single time point and investigated how well they correlate. Due to the lack of consensus in the field surrounding how best to define or assess flexibility, an obvious first step is to investigate whether self-report tests and neuropsychological assessments of flexibility that are commonly utilised in the field are well correlated. Identifying this gap will ideally lead to a better understanding of how best to define and/or assess flexibility. Findings from this review may allow for recommendations for the use of specific instruments in future research or clinical practice assessing flexibility in patients, thus aiding the scientist-practitioner model. Additionally, this review will facilitate increased consistency in administering both self-report and neuropsychological tests and correlating performances in research, making it less challenging to interpret and compare results between

studies assessing flexibility, and ideally lead to a comprehensive understanding of the construct of flexibility.

4. Objectives

The primary objective of this study is to systematically review the literature for quantitative evidence to support a relationship between self-report and neuropsychological tests used to assess flexibility tests in adults. To this end, the proposed systematic review will aim to answer the following question: *What is the relationship between self-report and neuropsychological tests used to assess flexibility in adults?*

5. Methods

Inclusion criteria for study eligibility

Types of participants

Inclusion criteria: Human adults (>18 years), any sex and/or gender, with any condition (healthy, clinical, and sub-clinical diagnoses (e.g., depression) or comorbid diagnoses) will be included.

Exclusion criteria: Studies that include animals or participants <18 years will not be included.

Types of study designs

Inclusion criteria: Primary experimental or observational studies of any design will be included, but studies must report at least one self-report and at least one neuropsychological assessment. Studies must administer the assessments within 1 month of each other, with no intervention between. Relevant studies will be included regardless of language or publication date/status.

Exclusion criteria: Intervention studies will be excluded if the two forms of tests are not administered during the same time point (i.e., if the two tests are separated by the intervention). Studies in which the two assessments are administered on different participants (e.g., and assume equivalence between the samples) will be excluded. Studies that use psychoactive substances or neural stimulation at baseline will be excluded. Post-intervention data of studies using psychoactive substances or neural stimulation as the main intervention will be also excluded.

Types of exposures

Studies included in this review must evaluate the associations between at least one self-report assessment and one neuropsychological assessment of flexibility. A self-report assessment has been defined as any type of questionnaire that involves asking participants to report their feelings, attitudes, beliefs or behaviours to assess flexibility. A neuropsychological assessment has been defined as any performance-based battery and/or single test used to assess flexibility. Examples of self-report tests of flexibility include: The Cognitive Flexibility Scale; the Cognitive Flexibility Inventory; the Attributional Style Questionnaire; or and any other tests to which we are naïve, that are identified during the search that assess flexibility. Examples of neuropsychological tests of flexibility include: the Wisconsin Card Sorting Test (WCST); the Trail Making Test (TMT); the Stroop Colour and Word Test; the Alternate Uses Task; or any other tests identified during the search that assess flexibility. Additional tests identified during the search must be stated *a priori* by the authors of that paper as an assessment of flexibility.

Types of outcome measures

Primary outcomes

To be included, the studies will be required to report mean and variance data for at least one self-report and one neuropsychological test of flexibility, respectively (authors can also provide statistical data within a specified time frame). Where correlational statistics are reported, (Pearson's *r* for parametric tests or Spearman's rank correlation coefficient for non-parametric tests), and confidence intervals (or standard errors)), we will extract them. Where correlational statistics are not reported, we will contact the authors to request them, or the raw data from which we can calculate them.

6. Search strategy for identification of studies

Electronic database searches

The primary researcher will execute the search strategy and finalise the search strategy with the help of an academic librarian. Initially, a literature search strategy will be developed using Medical Subject Headings (MeSH) and keywords related to self-report and neuropsychological assessments of flexibility. The following electronic databases will be searched: Cochrane Central Register of Controlled Trials (CENTRAL); MEDLINE (via OvidSP); EMBASE (via OvidSP); Emcare (via OvidSP); SCOPUS (via Elsevier); PsycINFO (via OvidSP); Web of Science Core Collection; the Cochrane Database of Systematic Reviews; PEDro; OT seeker. Searches will be modified to suit the functionality of each database. These databases were chosen as they are not only health focused and/or cover a broad range of relevant literature within the field, but are recommended for optimal database combinations for literature searches[24]. To ensure literature saturation, reference lists of all relevant systematic reviews and studies for which full-texts are included for data extraction will be manually searched to identify any additional studies. Secondary data (e.g., systematic reviews or meta-analyses) will not be included in this review.

Searching other resources

In addition to primary and secondary databases, grey literature will be searched, including clinical trial registries (e.g., World Health Organization International Clinical Trials Registry Platform (WHO ICTRP), Australian New Zealand Clinical Trials Registry (ANZCTR), Clinicaltrials.gov)), dissertations (e.g., ProQuest, Trove), Google Scholar (first 100 references) and conference abstracts. All included articles retrieved from Google Scholar will be saved as a PDF file and a data repository will be made publicly available. Conference abstracts will initially be searched (by author and title) to retrieve the original paper, but if this is unsuccessful, the authors will be contacted to request a full-text version of the paper. If the data is unavailable and not reported in the paper, we will request the raw data. We will identify potential key experts in the field of flexibility in SCOPUS to identify any studies that may have been missed or studies awaiting publication. Other channels of communication (e.g., formal emails) will be utilised if necessary, to consult key experts or influential authors in the field. Grey literature will only be included in this review if the author can be contacted and data are available for analysis. Book chapters will be excluded.

7. Data collection and analysis

Selection of studies

Studies identified by the systematic search will be exported to Endnote (Clarivate Analytics, Philadelphia, PA, USA) where duplicates will be removed, and then uploaded to Covidence (Covidence.org). Two independent reviewers will screen the articles by title and abstract to identify potentially relevant studies. We will exclude studies that clearly do not meet the inclusion criteria. The full-texts of all studies that either meet the inclusion criteria or lacked sufficient information in the abstract to exclude them will be retrieved and formally evaluated to determine study eligibility as per the pre-determined inclusion/exclusion criteria. Where we are unable to obtain a full-text paper, we will contact the authors and give them 2 weeks to provide us with either the published or unpublished data. Data for those authors that do not respond within the 2-week time frame will be considered irretrievable and will not be included. Two independent reviewers will complete this process; any disagreements will be resolved through discussion, and a third reviewer will be sought if necessary (i.e., if consensus cannot be adequately reached). Non-English publications will be translated using Google Translate, and extracted data will then be checked with a fluent speaker of the original publication language.

Data extraction and management

Data extraction will be completed by two independent reviewers using a piloted, customised form; any disagreements will be resolved by discussion, or if required, by consulting a third reviewer if consensus cannot be reached. The following data will be extracted: Citation; study aim(s); study design; study location; language; participant inclusion/exclusion criteria; sample size; number of drop outs; participant characteristics (age, sex and/or gender, health status (if unhealthy, we will extract primary diagnosis)); type(s) of tool pairing used (e.g. Cognitive Flexibility Scale and Wisconsin Card Sorting Test) this will herein be referred to as an 'assessment pair'; length of time between administering tests; time point of data collection (no-intervention, pre-intervention, post-intervention); concomitant interventions; test reliability; definition of flexibility; outcome data of comparison (e.g. test statistics: *p*-value; Pearson's *r* correlation coefficient or Spearman's rank correlation coefficient; standard errors or 95% confidence intervals; mean and standard deviation of each tool utilised). Where multiple assessment pairs are reported for different time points in a single study, each will be considered a unique assessment pair (any impact on pooled statistics will be considered at the analysis stage).

Risk of bias assessment of included studies

Risk of bias of individual studies will be assessed using an adapted version of tools suitable used to critically appraise cross-sectional data. The adapted tool will be customised, such that each included study will be critically appraised according to specific criteria or *key domains* that are relevant to the review question of interest (with reasons documented for judging studies with a high risk of bias). Two reviewers will independently assess the quality and risk of bias of included studies, and disagreements will be resolved either through discussion or by involvement of a third reviewer. Two researchers will pilot the risk of bias assessment tools on a sample of articles to ensure inter-rater agreement of included studies using an iterative process (>75% agreement of items within an individual study).

Dealing with missing data

If data are not available in the published study, the study's authors will be contacted a maximum of 3 times (via email or telephone) over a period of six weeks to obtain the missing data or request that the authors share the raw data for further analysis. If the authors are unable to be contacted, or they are unable to provide the data, the data will be considered irretrievable and will not be included.

Assessment of heterogeneity

Statistical heterogeneity will be assessed using χ^2 test and I^2 statistic. In accordance with the Cochrane criteria, when χ^2 has $p < 0.10$, this will be interpreted as statistically significant heterogeneity and when $I^2 > 50\%$, this will be interpreted as substantial heterogeneity[25].

Data synthesis strategy

Included studies will be pooled according to assessment pairs and the sampled population (i.e., healthy vs. diagnostic condition(s)). If pooled coefficients are similar (i.e., mean correlation within 1 SD), a meta-analysis will be performed on the entire dataset (i.e., all assessment pairs will be considered equivalent) using a random effects model in RevMan 5.3 software where effect estimates will be presented as forest plots. According to Bishara and Hittner (2015)[26], non-normally distributed data can inflate the Pearson correlation coefficient by up to .14, particularly in the presence of outliers; using the Spearman rank correlation coefficient produced conservative estimates. Thus, although Spearman rank and Pearson *r* correlation coefficients describe data in different ways, are calculated differently, and are therefore not strictly interchangeable, they provide a distribution of values that are broadly similar. As such, we will not convert Spearman's to Pearson's. We will instead assume that where normality was violated, the authors used Spearman's instead of Pearson's and we will pool studies regardless of which statistics they use. Where studies include several assessment pairs, we will divide the sample size between the number of paired tools to reduce the impact of one study on the pooled data. If substantial heterogeneity is apparent in the pooled effect estimates, included studies will be qualitatively synthesised and a narrative review of findings will be conducted instead.

Analysis of subgroups or subtests

Exploratory *post hoc* analyses (where data are available) will be performed and pooled in the subgroup of age.

Sensitivity analysis

If sufficient data are available, a sensitivity analyses will be performed considering: (i) papers with and without a conflict of interest (i.e., declared financial interests associated with the utilised tools); (ii) papers where the correlation was the primary hypothesis versus papers where it was not; (iii) papers that registered an *a priori* protocol versus papers that did not. Additionally, we will undertake a sensitivity analysis on post-intervention data to ensure that any interventions did not selectively alter the self-report or behavioural data. To do this, we will repeat group analyses with post-intervention data removed.

8. Amendment history

Amendment number	Protocol number	Date of change	Author(s) of changes	Details of changes
1	Version 2	24/04/2019	C. Howlett	Removed: decided not to search the PubMed database as Medline is a subset of PubMed. See: https://www.nlm.nih.gov/bsd/difference.html
2	Version 2	02/08/2019	C. Howlett and GL. Moseley	Minor change: in instances where entire conference proceedings were retrieved during title and abstract screening and did not provide an individual abstract, key terms (i.e. "flexib", "wisconsin", "trail mak" and "stroop") were searched within the entire proceeding. Any relevant abstracts progressed to full-text screening.
3	Version 2	14/10/2019	C. Howlett and GL. Moseley	Major change: we deviated from the initial protocol by including papers that used the Behaviour Rating Inventory of Executive Function (BRIEF) - a self-report test that includes a *shifting* sub-scale. Despite being naïve to this test and thus deviating from the rule: "additional tests identified during the search must be stated a priori by the authors of that paper as an assessment of flexibility", we gained consensus from colleagues and a substantial portion of the literature reported that this test assessed flexibility.
4	Version 2	30/10/2019	C. Howlett and GL. Moseley	Added: a sensitivity analysis on post-intervention data.
5	Version 2	30/10/2019	C. Howlett and GL. Moseley	Major change: due to resources and time constraints (screening >700 full-texts) we have decided to manually search the reference lists of all papers that progress to *data extraction.*
6	Version 2	30/10/2019	C. Howlett and GL. Moseley	Major change: during the full-text screening phase of the systematic review it was brought to the primary reviewer's attention that papers that used, for example, the Acceptance and Action Questionnaire (AAQ) – a renowned test of flexibility – would not be eligible due to the rule "*additional tests identified during the search must be stated a priori by the authors of that paper as an assessment of flexibility".* This is because of the nature of the search strategy. When trialling the search strategy, the AAQ and other tests to which we were not naïve to were being detected in the search results when using the wider terms "psychological", "mental" and "cognitive" flexibility, and thus the primary author did not include such tests in the search strategy directly. However, we have decided not to exclude these tests, but we will continue to implement the rule above

				when the authors are indeed naïve to other tests of flexibility.
7	Version 2	25/11/2019	C. Howlett and GL. Moseley	Added: a 2-week time frame for authors to provide published or unpublished data where we are unable to retrieve a full-text paper.
8	Version 3	30/03/2020	C. Howlett and GL. Moseley	Major change: studies using psychoactive substances or neural stimulation at baseline will be excluded. We will also not include post-intervention data of studies that use psychoactive substances or neural stimulation as the main intervention.
9	Version 3	08/04/2020	C. Howlett and GL. Moseley	Minor change: test validity was removed from data extraction as assessing the validity of tests is impossible. The following items, initially part of data extraction, were instead considered as part of the risk of bias assessment: statistical analysis techniques, whether assumptions of normality have been verified (statistically or not); where those data are not reported we will ask the author to clarify, conflict(s) of interest and a priori registered protocol.
10	Version 4	27/08/2020	C. Howlett and GL. Moseley	Major change: the use of RevMan software (the Cochrane Collaboration, 2014); however, the program did not have the functionality required to perform such correlational meta-analyses. We decided to perform the meta-analyses in R (R Core Team, 2019), using the *metafor* and *esc* packages.
11	Version 4	08/09/2020	C. Howlett, C. Berryman and GL. Moseley.	Major change: a sensitivity analysis on post-intervention data. We decided to not perform the sensitivity analysis because we found little evidence to support the relationship between self-report and neuropsychological tests of cognitive flexibility.
12	Version 4	23/10/2020	C. Howlett and GL. Moseley	Major change: we decided to only re-run the searches for the databases and not grey literature.
13	Version 4	26/10/2020	C. Howlett and GL. Moseley	Major change: we decided to not report the sensitivity analysis because our findings did not yield strong evidence to suggest that a relationship between self-report and neuropsychological tests of cognitive flexibility exists.

14	Version 4	03/11/2020	C. Howlett and GL. Moseley	Major change: Our initial search strategy and request for raw or aggregate data was very time intensive and onerous for the authors of eligible papers. It is for the aforementioned reasons that for our updated search strategy we decided to only include additional studies if the data were reported in the published paper. In other words, we did not request raw or aggregate data during this stage of the review.

References

1 Henschke N, Kamper S., Maher C. The epidemiology and economic consequences of pain. In: Mayo Clinic Proceedinga. Elsevier, 2015: 139–47.

2 Elliott AM, Smith BH, Penny KI, Cairns S. The epidemiology of chronic pain in the community. *Lancet* 1999; **354**: 1248–52.

3 Access Economics. The high price of pain: The economic impact of persistent pain in Australia. 2007 https://www.painaustralia.org.au/static/uploads/files/mbf-economic-impact-wffhrlzqsaht.pdf.

4 Gaskin DJ, Richard P. The economic costs of pain in the United States. *J Pain* 2012; **13**: 715–24.

5 Diamond A. Executive functions. *Annu Rev Psychol* 2013; **64**: 135–68.

6 Dajani DR, Uddin LQ. Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends Neurosci* 2015; **38**: 571–8.

7 Ionescu T. Exploring the nature of cognitive flexibility. *New Ideas Psychol* 2012; **30**: 190–200.

8 Dennis JP, Vander Wal JS. The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. *Cognit Ther Res* 2010; **34**: 241–53.

9 Johnco C, Wuthrich V, Rapee R. Reliability and validity of two self-report measures of cognitive flexibility. *Psychol Assess* 2014; **26**: 1381–7.

10 Lounes N, Khan G, Tchanturia K. Assessment of cognitive flexibility in anorexia nervosa-self-report or experimental measure? A brief report. *J Int Neuropsychol Soc* 2011; **17**: 925–8.

11 Martin MM, Rubin RB. A new measure of cognitive flexibility. *Psychol Rep* 1995; **76**: 623–6.

12 Peterson C, Semmel A, von Baeyer C, Abramson LY, Metalsky GI, Seligman MEP. The Attributional Style Questionnaire. *Cognit Ther Res* 1982; **6**: 287–99.

13 Golden CJ. A group version of the Stroop Color and Word Test. *J Pers Assess* 1975; **39**: 386–8.

14 Reitan RM, Wolfson D. The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation. *Reitan Neuropsychol* 1985.

15 Berg E. A simple objective technique for measuring flexibility in thinking. *J Gen Psychol* 1948; **39**: 15–22.

16 Wilson RC, Guilford JP, Christensen PR, Lewis DJ. A factor analytic study of creative-thinking abilities. *Psychometrika* 1954; **19**: 297–311.

17 Lee R, Hermens D, Porter M, Redoblado-Hodge A. A meta-analysis of cognitive deficits in first-episode major depressive disorder. *J Affect Disord* 2012; **140**: 113–24.

18 Meiran N, Diamond GM, Toder D, Nemets B. Cognitive rigidity in unipolar depression and obsessive compulsive disorder: Examination of task switching, Stroop, working memory updating and post-conflict adaptation. *Psychiatry Res* 2011; **185**: 149–56.

19 Lee J, Orsillo S. Investigating cognitive flexibility as a potential mechanism of mindfulness in generalized anxiety disorder. *J Behav Ther Exp Psychiatry* 2014; **45**: 208–16.

20 Tchanturia K, Davies H, Roberts M, *et al.* Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PLoS One* 2012; **7**: e28331.

21 Roberts ME, Tchanturia K, Stahl D, Southgate L, Treasure J. A systematic review and meta-analysis of set-shifting ability in eating disorders. *Psychol Med* 2007; **37**: 1075–84.

22 Berryman C, Stanton TR, Bowering KJ, Tabor A, McFarlane A, Moseley GL. Do people with chronic pain have impaired executive function? A meta-analytical review. *Clin Psychol Rev* 2014; **34**: 563–79.

23 Attal N, Masselin-Dubois A, Martinez V, *et al.* Does cognitive functioning predict chronic pain? Results from a prospective surgical cohort. *Brain* 2014; **137**: 904–17.

24 Bramer WM, Rethlefsen ML, Kleijnen J, Franco OH. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *2Systematic Rev* 2017; **6**. DOI:10.1186/s13643-017-0644-y.

25 Higgins J, Green S. Cochrane hanbook for systematic reviews of interventions. The Cochrane Collaboration, 2011 http://handbook.cochrane.org.

26 Bishara AJ, Hittner JB. Reducing bias and error in the correlation coefficient due to nonnormality. *Educ Psychol Meas* 2015; **75**: 785–804.

Supplementary 2 (S2): Search strategy for Chapters Three and Four

Black literature databases (from inception to 14th May 2019).

Medline search strategy.

Database: Ovid MEDLINE (R)
Search Strategy:
1. stroop test/ (1526) 2. trail making test/ (770) 3. wisconsin card sorting test/ (25) 4. (brixton task or brixton spatial anticipation task or catbat task or haptic illusion task or haptic illusion* or "set?shifting subset of cantab" or cantab or cambridge neuropsychological test? or cambridge neuropsychological test automated battery or intra?extra dimensional set shift* or intra dimensional shift* test or extra dimensional shift* test or IDED or IED or multi?task* test? or MTT or delis-kaplan executive function* or delis kaplan executive function* or stroop or colo?r?word interference test? or halstead* or seashore rhythm? or m-wcst or wisconsin card? sort* or modified wisconsin card sort* test or alternat* use? task? or trail mak* or verbal fluency test or word context test).mp. (64925) 5. (task?switching or task-switching or mental* flexib* or cognitive* flexib* or psychological* flexib* or cognitive* inflexib* or mental* inflexib* or psychological* inflexib* or set?shifting or set?shift*).mp. (5052) 6. 1 or 2 or 3 or 4 or 5 (69322) 7. self report/ (26758) 8. (cognitive flexibility scale? or cognitive flexibility inventor* or attributional style question?aire? or "detail and flexibility question?aire" or DFlex or "coping style? and flexib* inventor*").mp. (140) 9. (self?report* or self report*).mp. (148801) 10. 7 or 8 or 9 (148925) 11. 6 and 10 (816)

Embase search strategy.

Database: Ovid Embase
Search Strategy:
1. stroop test/ (5543) 2. trail making test/ (2202) 3. wisconsin card sorting test/ (2930) 4. (brixton task or brixton spatial anticipation task or catbat task or haptic illusion task or haptic illusion* or "set?shifting subset of cantab" or cantab or cambridge neuropsychological test? or cambridge neuropsychological test automated battery or intra?extra dimensional set shift* or intra dimensional shift* test or extra dimensional shift* test or IDED or IED or multi?task* test? or MTT or delis-kaplan executive function* or delis kaplan executive function* or stroop or colo?r?word interference test? or halstead* or seashore rhythm? or m-wcst or wisconsin card? sort* or modified wisconsin card sort* test or alternat* use? task? or trail mak* or verbal fluency test or word context test).mp. (111954) 5. (task?switching or task-switching or mental* flexib* or cognitive* flexib* or

psychological* flexib* or cognitive* inflexib* or mental* inflexib* or psychological* inflexib* or set?shifting or set?shift*).mp. (6753)
6. 1 or 2 or 3 or 4 or 5 (117544)
7. self report/ (109210)
8. (cognitive flexibility scale? or cognitive flexibility inventor* or attributional style question?aire? or "detail and flexibility question?aire" or DFlex or "coping style? and flexib* inventor*").mp. (194)
9. (self?report* or self report*).mp. (212206)
10. 7 or 8 or 9 (212369)
11. 6 and 10 (1388)

Emcare search strategy.

Database: Ovid Emcare
Search Strategy:

1. stroop test/ (1584)
2. trail making test/ (723)
3. wisconsin card sorting test/ (968)
4. (brixton task or brixton spatial anticipation task or catbat task or haptic illusion task or haptic illusion* or "set?shifting subset of cantab" or cantab or cambridge neuropsychological test? or cambridge neuropsychological test automated battery or intra?extra dimensional set shift* or intra dimensional shift* test or extra dimensional shift* test or IDED or IED or multi?task* test? or MTT or delis-kaplan executive function* or delis kaplan executive function* or stroop or colo?r?word interference test? or halstead* or seashore rhythm? or m-wcst or wisconsin card? sort* or modified wisconsin card sort* test or alternat* use? task? or trail mak* or verbal fluency test or word context test).mp. (15603)
5. (task?switching or task-switching or mental* flexib* or cognitive* flexib* or psychological* flexib* or cognitive* inflexib* or mental* inflexib* or psychological* inflexib* or set?shifting or set?shift*).mp. (1955)
6. 1 or 2 or 3 or 4 or 5 (17249)
7. self report/ (61112)
8. (cognitive flexibility scale? or cognitive flexibility inventor* or attributional style question?aire? or "detail and flexibility question?aire" or DFlex or "coping style? and flexib* inventor*").mp. (87)
9. (self?report* or self report*).mp. (97193)
10. 7 or 8 or 9 (97266)
11. 6 and 10 (545)

PsycINFO search strategy.

Database: Ovid PsycINFO
Search Strategy:

1. Stroop Color Word Test/ (680)
2. Halstead Reitan Neuropsychological Battery/ (170)
3. Wisconsin Card Sorting Test/ (441)
4. (brixton task or brixton spatial anticipation task or catbat task or haptic illusion task or haptic illusion* or "set?shifting subset of cantab" or cantab or cambridge neuropsychological test? or cambridge neuropsychological test automated battery or intra?extra dimensional set shift* or intra dimensional shift* test or extra dimensional shift* test or IDED or IED or multi?task* test? or MTT or delis-kaplan executive function* or delis kaplan executive function* or stroop or colo?r?word interference test? or halstead* or seashore rhythm? or m-wcst or wisconsin card? sort* or modified wisconsin card sort* test or alternat* use? task? or trail mak* or verbal fluency test or word context test).mp. (33852)
5. (task?switching or task-switching or mental* flexib* or cognitive* flexib* or psychological* flexib* or cognitive* inflexib* or mental* inflexib* or psychological* inflexib* or set?shifting or set?shift*).mp. (6730)
6. 1 or 2 or 3 or 4 or 5 (39157)
7. self-report/ (15892)
8. (cognitive flexibility scale? or cognitive flexibility inventor* or attributional style question?aire? or "detail and flexibility question?aire" or DFlex or "coping style? and flexib* inventor*").mp. (1056)
9. (self?report* or self report*).mp. (127021)
10. 7 or 8 or 9 (127931)
11. 6 and 10 (2313)

Cochrane Library search strategy.

Database: Cochrane Library	
Search Strategy:	
#1	MeSH descriptor: [Stroop Test] this term only (121)
#2	MeSH descriptor: [Trail Making Test] this term only (56)
#3	MeSH descriptor: [Wisconsin Card Sorting Test] this term only (1)
#4	("brixton task" or "brixton spatial anticipation task" or "catbat task" or "haptic illusion task" or haptic illusion* or set?shifting NEXT subset NEXT of NEXT cantab or cantab or cambridge NEXT neuropsychological NEXT test? or "cambridge neuropsychological test automated battery" or intra?extra NEXT dimensional NEXT set NEXT shift* or intra NEXT dimensional NEXT shift* NEXT test or extra NEXT dimensional NEXT shift* NEXT test or IDED or IED or multi?task* NEXT test? or MTT or delis-kaplan NEXT executive NEXT function* or delis NEXT kaplan NEXT executive NEXT function* or stroop or colo?r?word NEXT interference NEXT test? or halstead* or seashore NEXT rhythm? or m-wcst or wisconsin NEXT card? NEXT sort* or modified NEXT wisconsin NEXT card NEXT sort* test or alternat* NEXT use? NEXT task? or trail NEXT mak* or "verbal fluency test" or "word context test") (3839)
#5	(task?switching or task-switching or mental* NEXT flexib* or cognitive* NEXT flexib* or psychological* NEXT flexib* or cognitive* NEXT inflexib* or mental* NEXT inflexib* or psychological* NEXT inflexib* or set?shifting or set?shift*) (5)

#6	#1 or #2 or #3 or #4 or #5 (3843)
#7	MeSH descriptor: [Self Report] this term only (1934)
#8	(cognitive NEXT flexibility NEXT scale? or cognitive NEXT flexibility NEXT inventor* or attributional NEXT style NEXT question?aire or "detail and flexibility" NEXT question?aire or DFlex or "coping style? and" NEXT flexib* NEXT inventor*) (35)
#9	(self?report* or self NEXT report*) (8705)
#10	#7 or #8 or #9 (10209)
#11	#6 and #10 (119)

Web of Science search strategy.

Database: Clarivate Analytics Web of Science	
Search Strategy:	
#1	TS=("brixton task" or "brixton spatial anticipation task" or "catbat task" or "haptic illusion task" or "haptic illusion*" or "set?shifting subset of cantab" or cantab or "cambridge neuropsychological test?" or "cambridge neuropsychological test automated battery" or "intra?extra dimensional set shift*" or "intra dimensional shift* test" or "extra dimensional shift* test" or IDED or IED or "multi?task* test? " or MTT or "delis-kaplan executive function*" or "delis kaplan executive function*" or stroop or "colo?r?word interference test?" or halstead* or "seashore rhythm?" or "m-wcst" or "wisconsin card? sort*" or "modified wisconsin card sort* test" or "alternat* use? task?" or "trail mak*" or "verbal fluency test" or "word context test" or task?switching or "task-switching" or "mental* flexib*" or "cognitive* flexib*" or "psychological* flexib*" or "cognitive* inflexib*" or "mental* inflexib*" or "psychological* inflexib*" or set?shifting or set?shift*) *DocType=All document types; Language=All languages;* *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
#2	TS=(self?report* or "self report*" or "cognitive flexibility scale?" or "cognitive flexibility inventor*" or "attributional style question?aire" or "detail and flexibility question?aire" or DFlex or "coping style? and flexib* inventor*") *DocType=All document types; Language=All languages;* *Indexes=SCI-EXPANDED, SSCI Timespan=All years*
#3	#2 AND #1 *DocType=All document types; Language=All languages;* *Indexes=SCI-EXPANDED, SSCI Timespan=All years* Total: 883 records

SCOPUS search strategy.

Database: SCOPUS
Search Strategy:
(((TITLE-ABS-KEY ("brixton task" OR "brixton spatial anticipation task" OR "catbat task" OR "haptic illusion task" OR "haptic illusion*" OR "set?shifting subset of cantab" OR cantab OR "cambridge neuropsychological test?") OR TITLE-ABS-KEY ("cambridge neuropsychological test automated battery" OR "intra?extra dimensional set shift*" OR "intra dimensional shift* test" OR "extra dimensional shift* test" OR ided OR ied OR "multi?task* test?") OR TITLE-ABS-KEY (mtt OR "delis-kaplan executive function*" OR

"delis kaplan executive function*" OR stroop OR "colo?rword interference test?" OR halstead* OR "seashore rhythm?" OR "m-wcst" OR "wisconsin card? sort*") OR TITLE-ABS-KEY ("modified wisconsin card sort* test" OR "alternat* use? task?" OR "trail mak*") OR TITLE-ABS-KEY ("verbal fluency test" OR "word context test") OR TITLE-ABS-KEY (task?switching OR "task-switching" OR "mental* flexib*" OR "cognitive* flexib*" OR "psychological* flexib*") OR TITLE-ABS-KEY ("cognitive* inflexib*" OR "mental* inflexib*" OR "psychological* inflexib*" OR set?shifting OR set?shift*))) AND (TITLE-ABS-KEY (self?report* OR "self report*") OR TITLE-ABS-KEY ("cognitive flexibility scale?" OR "cognitive flexibility inventor*" OR "attributional style question?aire" OR "detail and flexibility question?aire" OR dflex OR "coping style? and flexib* inventor*")))

Total: 1142 records

OTseeker search strategy.

Database: OTseeker
Search Strategy:

Advanced Search: Any Field

Line 1 "brixton task" or "brixton spatial anticipation task" or "catbat task" or "haptic illusion task" or "haptic illusion*" or "set?shifting subset of cantab" or cantab or "cambridge neuropsychological test?" or "cambridge neuropsychological test automated battery" or "intra?extra dimensional set shift*" or "intra dimensional shift* test" or "extra dimensional shift* test" or IDED or IED or "multi?task* test?" or MTT or "delis-kaplan executive function*" or "delis kaplan executive function*" or stroop or "colo?r?word interference test?" or halstead* or "seashore rhythm?" or "m-wcst" or "wisconsin card? sort*" or "modified wisconsin card sort* test" or "alternat* use? task?" or "trail mak*" or "verbal fluency test" or "word context test" or task?switching or "task-switching" or "mental* flexib*" or "cognitive* flexib*" or "psychological* flexib*" or "cognitive* inflexib*" or "mental* inflexib*" or "psychological* inflexib*" or set?shifting or set?shift*

AND

Line 2 self?report* or "self report*" or "cognitive flexibility scale?" or "cognitive flexibility inventor*" or "attributional style question?aire" or "detail and flexibility question?aire" or DFlex or "coping style? and flexib* inventor*"

Total = 6 records

PEDro search strategy.

Database: PEDro
Search Strategy:
Abstract & Title: cognitive flexibility *Match all search terms (AND)* Total: 82 records

Grey literature databases

Google Scholar search strategy.

Database: Google Scholar (first 100 records)
Search Strategy:
(self-report) AND (neuropsychological) AND (cognitive flexibility OR mental flexibility OR psychological flexibility) Total: First 100 records

Database: Trove
Search Strategy:
("mental flexibility" OR "cognitive flexibility" OR "psychological flexibility" OR "wisconsin card sorting" OR "trail-making test" OR "trail making test") AND (self-report* OR "self report" OR "self reported" OR "cognitive flexibility scale" OR "cognitive flexibility inventory" OR "flexibility questionaire" OR "flexibility questionnaire") Total: 73 records [2 exported]

ProQuest Dissertation & Theses

Database: ProQuest Dissertation & Theses
Search Strategy:
noft("brixton task" OR "brixton spatial anticipation task" OR "catbat task" OR "haptic illusion task" OR "haptic illusion*" OR "set?shifting subset of cantab" OR cantab OR "cambridge neuropsychological test?" OR "cambridge neuropsychological test automated battery" OR "intra?extra dimensional set shift*" OR "intra dimensional shift* test" OR "extra dimensional shift* test" OR ided OR ied OR "multi?task* test?" OR mtt OR "delis-kaplan executive function*" OR "delis kaplan executive function*" OR stroop OR "colo?rword interference test?" OR halstead* OR "seashore rhythm?" OR "m-wcst" OR "wisconsin card? sort*" OR "modified wisconsin card sort* test" OR "alternat* use? task?" OR "trail mak*" OR "verbal fluency test" OR "word context test" OR

task?switching OR "task-switching" OR "mental* flexib*" OR "cognitive* flexib*" OR "psychological* flexib*" OR "cognitive* inflexib*" OR "mental* inflexib*" OR "psychological* inflexib*" OR set?shifting OR set?shift*) AND noft(self?report* OR "self report*" OR "cognitive flexibility scale?" OR "cognitive flexibility inventor*" OR "attributional style question?aire" OR "detail and flexibility question?aire" OR dflex OR "coping style? and flexib* inventor*")

Total: 366 records [32 exported]

WHO International Clinical Trials Registry Platform search strategy.

Database: WHO International Clinical Trials Registry Platform (includes Australian New Zealand Clinical Trials Registry (ANZCTR) & Clinicaltrials.gov)) <last data file imported on July 29, 2019>
Search Strategy:
cognitive flexibility OR mental flexibility OR psychological flexibility

Total: 94 records [10 exported]

Australian New Zealand Clinical Trials Registry (ANZCTR) search strategy.

Database: ANZCTR <from July 30, 2019 to August 29, 2019>
Search Strategy:
cognitive flexibility OR mental flexibility OR psychological flexibility

Total: 0 records

Clinicaltrials.gov search strategy.

Database: Clinicaltrials.gov
Search Strategy:
cognitive flexibility OR mental flexibility OR psychological flexibility

Total: 0 records

Appendix B: Supplementary material for Chapter Three

Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, 88(102061). https://doi.org/10.1016/j.cpr.2021.102061

Outline:

- **Supplementary 1 (S1):** Full details on attempts to obtain data
- **Supplementary 2 (S2):** Meta-analyses of additional test pairs

Supplementary 1 (S1): Full details on attempts to obtain data

Reason for exclusion	**Citation(s) & year**
Author(s) who did not respond when contacted (*n* = 38)	Albein-Urios, Youssef, Kirkovski, & Enticott, 2018; Ballantyne, 2013; Baniqued et al., 2014; Bannon, 2017; Banz, 2014; Biederman et al., 2017; Boschloo, Krabbendam, Aben, De Groot, & Jolles, 2014; Coleman, 2010; Daly, Hildenbrand, Turner, Berkowitz, & Tarazi, 2017; Dickson, Ciesla, & Zelic, 2017; Dumbrava, 2017; Follmer & Stefanou, 2014; Fujino et al., 2017; García-Molina, Tormos, Bernabeu, Junqué, & Roig-Rovira, 2012; García, Torrecillas, Arcos, & García, 2005; Gelonch, Garolera, Valls, Rosselló, & Pifarré, 2016; Gonzalez, Figueroa, Bellows, Rhodes, & Youmans, 2013; Good, 2014; Hagen et al., 2016; Healy, Treadwell, & Reagan, 2011; Inozemtseva, Pérez-Solís, Matute, & Juárez, 2016; Kesler et al., 2013; Kesler, Kent, & O'Hara, 2011; Lee, 1999; Li et al., 2013; Oswald, 2015; Pope, Ross, & Stavrinos, 2016; Renwick et al., 2015; Rouel, Raman, Hay, & Smith, 2016; Sakamoto et al., 2013; Solsnes, Skranes, Brubakk, & Lohaugen, 2014; Sternheim, van der Burgh, Berkhout, Dekker, & Ruiter, 2014; Suchy et al., 2016; Thompson, 2010; Van Noort, Kraus, Pfeiffer, Lehmkuhl, & Kappel, 2016; Wenner, Bianchi, Figueredo, Rushton, & Jacobs, 2013; Wiebe, Baker, Suchy, Stump, & Berg, 2018; Wilens et al., 2017
Contact details could not be found (*n* = 4)	Kalkut, 2011; Nordvall et al., 2017; Salman, 1999; Tamati et al., 2013
Unable to provide data (*n* = 21)	Bettis et al., 2017; Demetriou et al., 2018; Garcia-Willingham, Roach, Kasarskis, & Segerstrom, 2018; Hanssen, Beiske, Landrø, & Hessen, 2014; Heinonen et al., 2013; Hocking et al., 2011; Lang et al., 2015; Lao-Kaim et al., 2015; Leppanen, Adamson, & Tchanturia, 2018; Løvstad, Funderud, Endestad, et al., 2012; Løvstad, Funderud, Meling, et al., 2012; Mangum, 2018; McAnarney et al., 2011; Mckowen et al., 2017; McKowen et al., 2018; Palm, 2005; Perkins, Cortina, Smith-Darden, & Graham-Bermann, 2012; Shuster & Toplak, 2009; Svendsen et al., 2017; Tornås et al., 2016; Westwood, Mandy, & Tchanturia, 2017
Article did not include a healthy cohort (*n* = 5)	Baker, Gibson, Georgiou-Karistianis, & Giummarra, 2018; Baker et al., 2018; Marceau, Berry, Lunn, Kelly, & Solowij, 2017; Okuda et al., 2017; Whiting et al., 2015
Author(s) were unable to determine which participants were using psychotropic medications at the time of testing (*n* = 1)	Donders et al., 2015
Duplicate data (*n* = 4)	Johnco et al., 2014; Marceau et al., 2018; McCurdy et al., 2016; Perkins et al., 2014
Participants ≥18 years did not complete self-report tests of cognitive flexibility (*n* = 1)	van Zellem et al., 2015
Author(s) were unable to confirm whether the assessment pair was completed within 1 month (*n* = 1)	Warren et al., 2013

References

Albein-Urios, N., Youssef, G. J., Kirkovski, M., & Enticott, P. G. (2018). Autism spectrum traits linked with reduced performance on self-report behavioural measures of cognitive flexibility. *Journal of Autism and Developmental Disorders*, *48*(7), 2506–2515. https://doi.org/10.1007/s10803-018-3503-3

Baker, K.S., Gibson, S. J., Georgiou-Karistianis, N., & Giummarra, M. J. (2018). Relationship between self-reported cognitive difficulties, objective neuropsychological test performance and psychological distress in chronic pain. *European Journal of Pain*, *22*(3), 601–613. https://doi.org/10.1002/ejp.1151

Baker, K. S., Georgiou-Karistianis, N., Lampit, A., Valenzuela, M., Gibson, S. J., & Giummarra, M. J. (2018). Computerised training improves cognitive performance in chronic pain: A participant-blinded randomised active-controlled trial with remote supervision. *Pain*, *159*(4), 644–655. https://doi.org/10.1097/j.pain.0000000000001150

Ballantyne, E. C. (2013). *The neuropsychology of mindfulness: Attentional aspects* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (3578540)

Baniqued, P. L., Kranz, M. B., Voss, M. W., Lee, H., Cosman, J. D., Severson, J., & Kramer, A. F. (2014). Cognitive training with casual video games: Points to consider. *Frontiers in Psychology*, *4*(1010), 1–19. https://doi.org/10.3389/fpsyg.2013.01010

Bannon, E. E. (2017). *The efficacy of acceptance based behavior therapy versus cognitive therapy for test anxiety and working memory performance* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (10754246)

Banz, B. C. (2014). *An evaluation of executive functions, cognitive control and a neurocognitive profile of college binge* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (3635581)

Bettis, A. H., Coiro, M. J., England, J., Murphy, L. K., Zelkowitz, R. L., Dejardins, L., Eskridge, R., Adery, L. H., Yarboi, J., Pardo, D., & Compas, B. E. (2017). Comparison of two approaches to prevention of mental health problems in college students: Enhancing coping and executive function skills. *Journal of American College Health*, *65*(5), 313–322. https://doi.org/10.1080/07448481.2017.1312411

Biederman, J., Fried, R., Tarko, L., Surman, C., Spencer, T., Pope, A., Grossman, R., McDermott, K.,

Woodworth, K. Y., & Faraone, S. V. (2017). Memantine in the treatment of executive function deficits in adults with ADHD: A pilot-randomized double-blind controlled clinical trial. *Journal of Attention Disorders*, *21*(4), 343–352. https://doi.org/10.1177/1087054714538656

Boschloo, A., Krabbendam, L., Aben, A., De Groot, R., & Jolles, J. (2014). Sorting Test, Tower Test, and BRIEF-SR do not predict school performance of healthy adolescents in preuniversity education. *Frontiers in Psychology*, *5*(287), 1–8. https://doi.org/10.3389/fpsyg.2014.00287

Coleman, S. L. (2010). *The relationship between cognitive inhibition and intrusive thoughts* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (3423272)

Daly, B. P., Hildenbrand, A. K., Turner, E., Berkowitz, S., & Tarazi, R. A. (2017). Executive functioning among college students with and without history of childhood maltreatment. *Journal of Aggression, Maltreatment and Trauma*, *26*(7), 717–735. https://doi.org/10.1080/10926771.2017.1317685

Demetriou, E. A., Song, C. Y., Park, S. H., Pepper, K. L., Naismith, S. L., Hermens, D. F., Hickie, I. B., Thomas, E. E., Norton, A., White, D., & Guastella, A. J. (2018). Autism, early psychosis, and social anxiety disorder: A transdiagnostic examination of executive function cognitive circuitry and contribution to disability. *Translational Psychiatry*, *8*(200), 1–10. https://doi.org/10.1038/s41398-018-0193-8

Dickson, K. S., Ciesla, J. A., & Zelic, K. (2017). The role of executive functioning in adolescent rumination and depression. *Cognitive Therapy and Research*, *41*, 62–72. https://doi.org/10.1007/s10608-016-9802-0

Donders, J., Oh, Y. I., & Gable, J. (2015). Self-and informant ratings of executive functioning after mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *30*(6), e30–e39. https://doi.org/10.1097/HTR.0000000000000120

Dumbrava, I. L. (2017). An empirical approach to wisdom processes. *Romanian Journal of Applied Psychology*, *19*(2), 50–58. https://doi.org/10.24913/rjap.19.2.04

Follmer, D. J., & Stefanou, C. R. (2014). Examining the correspondence between a direct and an indirect measure of executive functions: Implications for school-based assessment. *The School Psychologist*, *68*(3), 12–18.

Fujino, J., Tei, S., Jankowski, K. F., Kawada, R., Murai, T., & Takahashi, H. (2017). Role of spontaneous brain activity in explicit and implicit aspects of cognitive flexibility under

socially conflicting situations: A resting-state fMRI study using fractional amplitude of low-frequency fluctuations. *Neuroscience*, *367*, 60–71. https://doi.org/10.1016/j.neuroscience.2017.10.025

García-Molina, A., Tormos, J. M., Bernabeu, M., Junqué, C., & Roig-Rovira, T. (2012). Do traditional executive measures tell us anything about daily-life functioning after traumatic brain injury in Spanish-speaking individuals? *Brain Injury*, *26*(6), 864–874. https://doi.org/10.3109/02699052.2012.655362

Garcia-Willingham, N. E., Roach, A. R., Kasarskis, E. J., & Segerstrom, S. C. (2018). Self-regulation and executive functioning as related to survival in Motor Neuron Disease: Preliminary findings. *Psychosomatic Medicine*, *80*(7), 665–672. https://doi.org/10.1097/PSY.0000000000000602

García, A. V., Torrecillas, F. L., Arcos, F. A. De, & García, M. P. (2005). Effects of executive impairments on maladaptive explanatory styles in substance abusers: Clinical implications. *Archives of Clinical Neuropsychology*, *20*(1), 67–80. https://doi.org/10.1016/j.acn.2004.03.002

Gelonch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and objective measures. *Comprehensive Psychiatry*, *66*, 113–122. https://doi.org/10.1016/j.comppsych.2016.01.002

Gonzalez, C. A., Figueroa, I. J., Bellows, B. G., Rhodes, D., & Youmans, R. J. (2013). A new behavioral measure of cognitive flexibility. *In International Conference on Engineering Psychology and Cognitive Ergonomics*, 297–306. https://doi.org/10.1007/978-3-642-39360-0-33

Good, D. (2014). Predicting real-time adaptive performance in a dynamic decision-making context. *Journal of Management and Organization*, *20*(6), 715–732. https://doi.org/10.1017/jmo.2014.54

Hagen, E., Erga, A. H., Hagen, K. P., Nesvåg, S. M., McKay, J. R., Lundervold, A. J., & Walderhaug, E. (2016). Assessment of executive function in patients with substance use disorder: A comparison of inventory- and performance-based assessment. *Journal of Substance Abuse Treatment*, *66*, 1–8. https://doi.org/10.1016/j.jsat.2016.02.010

Hanssen, K. T., Beiske, A. G., Landrø, N. I., & Hessen, E. (2014). Predictors of executive complaints and executive deficits in multiple sclerosis. *Acta Neurologica Scandinavica*, *129*(4), 234–242. https://doi.org/10.1111/ane.12177

Healy, B., Treadwell, A., & Reagan, M. (2011). Measures of RSA suppression, attentional control, and negative affect predict self-ratings of executive functions. *Journal of Psychophysiology*, *25*(4), 164–173. https://doi.org/10.1027/0269-8803/a000053

Heinonen, K., Pesonen, A. K., Lahti, J., Pyhälä, R., Strang-Karlsson, S., Hovi, P., Järvenpää, A. L., Eriksson, J. G., Andersson, S., Kajantie, E., & Raikkonen, K. (2013). Self- and parent-rated executive functioning in young adults with very low birth weight. *Pediatrics*, *131*(1), e243–e250. https://doi.org/10.1542/peds.2012-0839

Hocking, M. C., Hobbie, W. L., Deatrick, J. A., Lucas, M. S., Szabo, M. M., Volpe, E. M., & Barakat, L. P. (2011). Neurocognitive and family functioning and quality of life among young adult survivors of childhood brain tumors. *Clinical Neuropsychologist*, *25*(6), 942–962. https://doi.org/10.1080/13854046.2011.580284

Inozemtseva, O., Pérez-Solís, L., Matute, E., & Juárez, J. (2016). Differential improvement of executive functions during abstinence in cocaine-dependent patients: A longitudinal study. *Substance Use & Misuse*, *51*(11), 1428–1440. https://doi.org/10.1080/10826084.2016.1178293

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). The influence of cognitive flexibility on treatment outcome and cognitive restructuring skill acquisition during cognitive behavioural treatment for anxiety and depression in older adults: Results of a pilot study. *Behaviour Research and Therapy*, *57*, 55–64. https://doi.org/10.1016/j.brat.2014.04.005

Kalkut, E. (2010). *Neuropsychological assessment of executive functioning and its association with depressive symptomology* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (3419840)

Kesler, S., Hadi Hosseini, S. M., Heckler, C., Janelsins, M., Palesh, O., Mustian, K., & Morrow, G. (2013). Cognitive training for improving executive function in chemotherapy-treated breast cancer survivors. *Clinical Breast Cancer*, *13*(4), 299–306. https://doi.org/10.1016/j.clbc.2013.02.004

Kesler, S. R., Kent, J. S., & O'Hara, R. (2011). Prefrontal cortex and executive function impairments in primary breast cancer. *Archives of Neurology*, *68*(11), 1447–1453. https://doi.org/10.1001/archneurol.2011.245

Lang, K., Lloyd, S., Khondoker, M., Simic, M., Treasure, J., & Tchanturia, K. (2015). Do children and adolescents with anorexia nervosa display an inefficient cognitive processing style? *PLoS*

ONE, *10*(7), 1–13. https://doi.org/10.1371/journal.pone.0131724

Lao-Kaim, N. P., Fonville, L., Giampietro, V. P., Williams, S. C. R., Simmons, A., & Tchanturia, K. (2015). Aberrant function of learning and cognitive control networks underlie inefficient cognitive flexibility in anorexia nervosa: A cross-sectional fMRI study. *PLoS ONE*, *10*(5), 1–16. https://doi.org/10.1371/journal.pone.0124027

Lee, J. (1999). *Social intelligence: Flexibility and complexity are distinct from creativity* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (9924645)

Leppanen, J., Adamson, J., & Tchanturia, K. (2018). Impact of cognitive remediation therapy on neurocognitive processing in anorexia nervosa. *Frontiers in Psychiatry*, *9*(96), 1–5. https://doi.org/10.3389/fpsyt.2018.00096

Li, Y., Dong, F., Cao, F., Cui, N., Li, J., & Long, Z. (2013). Poly-victimization and executive functions in junior college students. *Scandinavian Journal of Psychology*, *54*(6), 485–492. https://doi.org/10.1111/sjop.12083

Løvstad, M., Funderud, I., Endestad, T., Due-Tonnessen, P., Meling, T. R., Lindgren, M., Knight, R. T., & Solbakk, A. K. (2012). Executive functions after orbital or lateral prefrontal lesions: Neuropsychological profiles and self-reported executive functions in everyday living. *Brain Injury*, *26*(13–14), 1586–1598. https://doi.org/10.3109/02699052.2012.698787

Løvstad, M., Funderud, I., Meling, T., Krämer, U. M., Voytek, B., Due-Tønnessen, P., Endestad, T., Lindgren, M., Knight, R. T., & Solbakk, A. K. (2012). Anterior cingulate cortex and cognitive control: Neuropsychological and electrophysiological findings in two patients with lesions to dorsomedial prefrontal cortex. *Brain and Cognition*, *80*(2), 237–249. https://doi.org/10.1016/j.bandc.2012.07.008

Mangum, R. W. (2018). *Self-understanding of executive functions in individuals with agenesis of the corpus callosum* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (10792935)

Marceau, E. M., Berry, J., Lunn, J., Kelly, P. J., & Solowij, N. (2017). Cognitive remediation improves executive functions, self-regulation and quality of life in residents of a substance use disorder therapeutic community. *Drug and Alcohol Dependence*, *178*, 150–158. https://doi.org/10.1016/j.drugalcdep.2017.04.023

Marceau, E. M., Kelly, P. J., & Solowij, N. (2018). The relationship between executive functions and

emotion regulation in females attending therapeutic community treatment for substance use disorder. *Drug and Alcohol Dependence*, *182*, 58–66. https://doi.org/10.1016/j.drugalcdep.2017.10.008

McAnarney, E. R., Zarcone, J., Singh, P., Michels, J., Welsh, S., Litteer, T., Wang, H., & Klein, J. D. (2011). Restrictive anorexia nervosa and set-shifting in adolescents: A biobehavioral interface. *Journal of Adolescent Health*, *49*(1), 99–101. https://doi.org/10.1016/j.jadohealth.2010.11.259

McCurdy, M. D., Turner, E. M., Barakat, L. P., Hobbie, W. L., Deatrick, J. A., Paltin, I., Fisher, M. J., & Hocking, M. C. (2016). Discrepancies among measures of executive functioning in a subsample of young adult survivors of childhood brain tumor: Associations with treatment intensity. *Journal of the International Neuropsychological Society*, *22*(9), 900–910. https://doi.org/10.1017/S1355617716000771

McKowen, J., Carrellas, N., Zulauf, C., Ward, E. N., Fried, R., & Wilens, T. (2017). Factors associated with attrition in substance using patients enrolled in an intensive outpatient program. *American Journal on Addictions*, *26*(8), 780–787. https://doi.org/10.1111/ajad.12619

McKowen, J. W., Isenberg, B. M., Carrellas, N. W., Zulauf, C. A., Ward, N. E., Fried, R. S., & Wilens, T. E. (2018). Neuropsychological changes in patients with substance use disorder after completion of a one month intensive outpatient treatment program. *American Journal on Addictions*, *27*(8), 632–638. https://doi.org/10.1111/ajad.12824

Nordvall, O., Jonsson, B., & Neely, A. S. (2017). Self-reported and performance-based measures of executive functions in interned youth. *Psychology, Crime and Law*, *23*(3), 240–253. https://doi.org/10.1080/1068316X.2016.1239725

Okuda, T., Asano, K., Numata, N., Hirano, Y., Yamamoto, T., Tanaka, M., Matsuzawa, D., Shimizu, E., Iyo, M., & Nakazato, M. (2017). Feasibility of cognitive remediation therapy for adults with autism spectrum disorders: A single-group pilot study. *Neuropsychiatric Disease and Treatment*, *13*, 2185–2191. https://doi.org/10.2147/NDT.S141555

Oswald, A. (2015). *Effects of repeated exposure to stressors on cognitive flexibility as a function of self-type and self-complexity: A multi-session experimental test* (Unpublished Doctoral dissertation). University of Newcastle, Sydney, Australia.

Palm, K. M. (2004). *An examination of mindfulness: Assessment and relationship to PTSD* (Doctoral

dissertation). Retrieved from ProQuest Dissertations and Theses (3161689)

Perkins, S. C., Cortina, K. S., Smith-Darden, J. P., & Graham-Bermann, S. A. (2012). The mediating role of self-regulation between intrafamilial violence and mental health adjustment in incarcerated male adolescents. *Journal of Interpersonal Violence*, *27*(7), 1199–1224. https://doi.org/10.1177/0886260511425240

Perkins, S. C., Smith-Darden, J., Ametrano, R. M., & Graham-Bermann, S. (2014). Typologies of violence exposure and cognitive processing in incarcerated male adolescents. *Journal of Family Violence*, *29*(4), 439–451. https://doi.org/10.1007/s10896-014-9600-y

Pope, C. N., Ross, L. A., & Stavrinos, D. (2016). Association between executive function and problematic adolescent driving. *Journal of Developmental and Behavioral Pediatrics*, *37*(9), 702–711. https://doi.org/10.1097/DBP.0000000000000353

Renwick, B., Musiat, P., Lose, A., Dejong, H., Broadbent, H., Kenyon, M., Loomes, R., Watson, C., Ghelani, S., Serpell, L., Richards, L., Johnson-Sabine, E., Boughton, N., Treasure, J., & Schmidt, U. (2015). Neuro- and social-cognitive clustering highlights distinct profiles in adults with anorexia nervosa. *International Journal of Eating Disorders*, *48*(1), 26–34. https://doi.org/10.1002/eat.22366

Rouel, M., Raman, J., Hay, P., & Smith, E. (2016). Validation of the Behaviour Rating Inventory of Executive Function-Adult Version (BRIEF-A) in the obese with and without binge eating disorder. *Eating Behaviors*, *23*, 58–65. https://doi.org/10.1016/j.eatbeh.2016.07.010

Sakamoto, M., Woods, S. P., Kolessar, M., Kriz, D., Anderson, J. R., Olavarria, H., Sasaki, A. W., Chang, M., Flora, K. D., Loftis, J. M., & Huckans, M. (2013). Protective effects of higher cognitive reserve for neuropsychological and daily functioning among individuals infected with hepatitis C. *Journal of NeuroVirology*, *19*(5), 442–451. https://doi.org/10.1007/s13365-013-0196-4

Salmon, D. (1999). *Attentional control and the deconstruction of the schemas underlying the perception of pain* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (9934176)

Shuster, J., & Toplak, M. E. (2009). Executive and motivational inhibition: Associations with self-report measures related to inhibition. *Consciousness and Cognition*, *18*(2), 471–480. https://doi.org/10.1016/j.concog.2009.01.004

Solsnes, A. E., Skranes, J., Brubakk, A. M., & Lohaugen, G. C. C. (2014). Executive functions in very-low-birth-weight young adults: A comparison between self-report and neuropsychological test results. *Journal of the International Neuropsychological Society*, *20*(5), 506–515. https://doi.org/10.1017/S1355617714000332

Sternheim, L., van der Burgh, M., Berkhout, L. J., Dekker, M. R., & Ruiter, C. (2014). Poor cognitive flexibility, and the experience thereof, in a subclinical sample of female students with obsessive-compulsive symptoms. *Scandinavian Journal of Psychology*, *55*(6), 573–577. https://doi.org/10.1111/sjop.12163

Suchy, Y., Turner, S. L., Queen, T. L., Durracio, K., Wiebe, D. J., Butner, J., Franchow, E. I., White, P. C., Murray, M. A., Swinyard, M., & Berg, C. A. (2016). The relation of questionnaire and performance-based measures of executive functioning with type 1 diabetes outcomes among late adolescents. *Health Psychology*, *35*(7), 661– 669. https://doi.org/10.1037/hea0000326

Svendsen, T. S., Erga, A. H., Hagen, E., McKay, J. R., Njå, A. L. M., Årstad, J., & Nesvåg, S. (2017). How to maintain high retention rates in long-term research on addiction: A case report. *Journal of Social Work Practice in the Addictions*, *17*(4), 374–387. https://doi.org/10.1080/1533256X.2017.1361831

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: A first report. *Journal of the American Academy of Audiology*, *24*(7), 616–634. https://doi.org/10.3766/jaaa.24.7.10

Thompson, R. (2010). *Understanding bulimia nervosa from a neuropsychological perspective: Impulsivity and binge-purge behavior in adolescent and young adult women* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (3372002)

Tornås, S., Lovstad, M., Solbakk, A. K., Evans, J., Endestad, T., Hol, P. K., Schanke, A. K., & Stubberud, J. (2016). Rehabilitation of executive functions in patients with chronic acquired brain injury with goal management training, external cuing, and emotional regulation: A randomized controlled trial. *Journal of the International Neuropsychological Society*, *22*(4), 436–452. https://doi.org/10.1017/S1355617715001344

van Noort, B. M., Kraus, M. K. A., Pfeiffer, E., Lehmkuhl, U., & Kappel, V. (2016). Neuropsychological and behavioural short-term effects of cognitive remediation therapy in adolescent anorexia nervosa: A pilot study. *European Eating Disorders Review*, *24*(1), 69–74. https://doi.org/10.1002/erv.2383

van Zellem, L., Buysse, C., Madderom, M., Legerstee, J. S., Aarsen, F., Tibboel, D., & Utens, E. M. (2015). Long-term neuropsychological outcomes in children and adolescents after cardiac arrest. *Intensive Care Medicine*, *41*(6), 1057–1066. https://doi.org/10.1007/s00134-015-3789-y

Warren, S. L., Crocker, L. D., Spielberg, J. M., Engels, A. S., Banich, M. T., Sutton, B. P., Miller, G. A., & Heller, W. (2013). Cortical organization of inhibition-related functions and modulation by psychopathology. *Frontiers in Human Neuroscience*, *7*(271), 1–13. https://doi.org/10.3389/fnhum.2013.00271

Wenner, C. J., Bianchi, J., Figueredo, A. J., Rushton, J. P., & Jacobs, W. J. (2013). Life History theory and social deviance: The mediating role of executive function. *Intelligence*, *41*(2), 102–113. https://doi.org/10.1016/j.intell.2012.11.004

Westwood, H., Mandy, W., & Tchanturia, K. (2017). The association between symptoms of autism and neuropsychological performance in females with anorexia nervosa. *Psychiatry Research*, *258*, 531–537. https://doi.org/10.1016/j.psychres.2017.09.005

Whiting, D. L., Deane, F. P., Ciarrochi, J., Mcleod, H. J., Simpson, G. K. (2015). Validating measures of psychological flexibility in a population with validating measures of psychological flexibility in a population with acquired brain injury acquired brain injury. *Psychological Assessment*, *27*(2), 415–423. https://doi.org/10.1037/pas0000050

Wiebe, D. J., Baker, A. C., Suchy, Y., Stump, T. K., & Berg, C. A. (2018). Individual differences and day-to-day fluctuations in goal planning and type 1 diabetes management. *Health Psychology*, *37*(7), 638–646. https://doi.org/10.1037/hea0000624

Wilens, T. E., Carrellas, N. W., Martelon, M. K., Yule, A. M., Fried, R., Anselmo, R., & McCabe, S. E. (2017). Neuropsychological functioning in college students who misuse prescription stimulants. *American Journal on Addictions*, *26*(4), 379–387. https://doi.org/10.1111/ajad.12551

Supplementary 2 (S2): Meta-analyses of additional test pairs

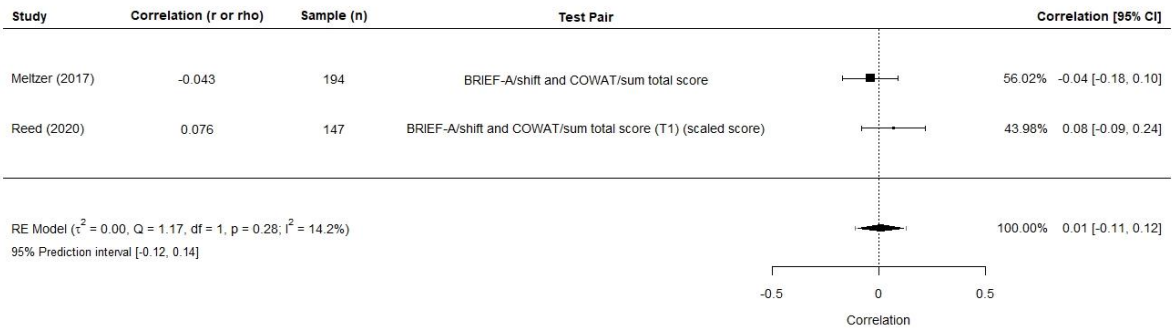


Figure S2.1. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the COWAT (total score).

A meta-analysis of two studies revealed that the BRIEF-A shift subscale (total score) and the COWAT (total score) did not correlate. That is, the test pair did not share a relationship (*r* = 0.01, 95% CI = -0.11 to 0.12).

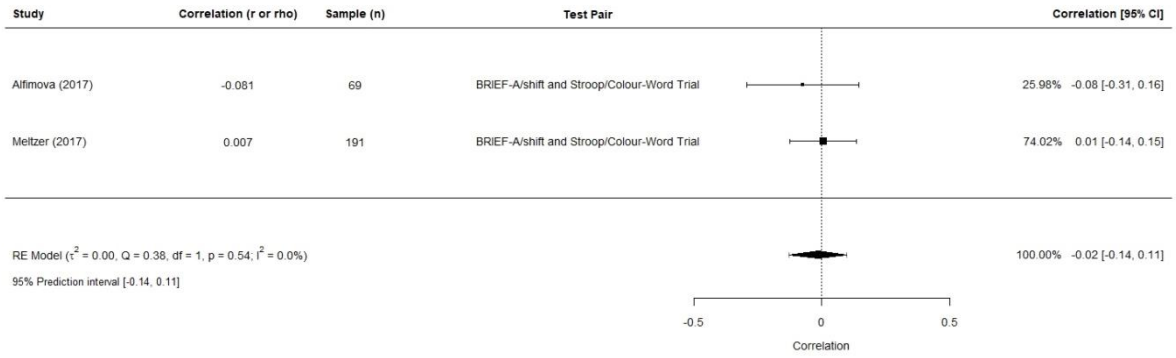


Figure S2.2. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the Stroop (Colour-Word trial).

A meta-analysis of two studies revealed that the BRIEF-A shift subscale (total score) and the Stroop (Colour-Word trial) did not relate. That is, the test pair did not share a relationship (*r* = -0.02, 95% CI = -0.14 to 0.11).

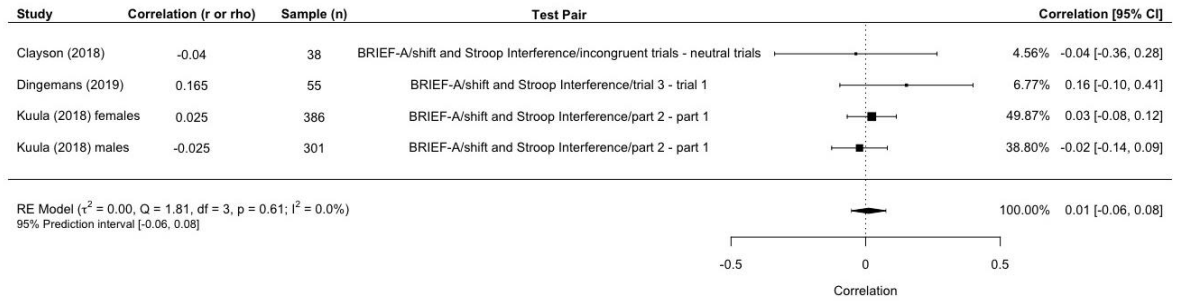


Figure S2.3. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the Stroop (interference score).

A meta-analysis of three studies (four cohorts) indicated that the BRIEF-A shift subscale (total score) and the Stroop (interference score) did not relate. That is, there was a non-significant effect estimate of 0.01 (95% CI = -0.06 to 0.08).

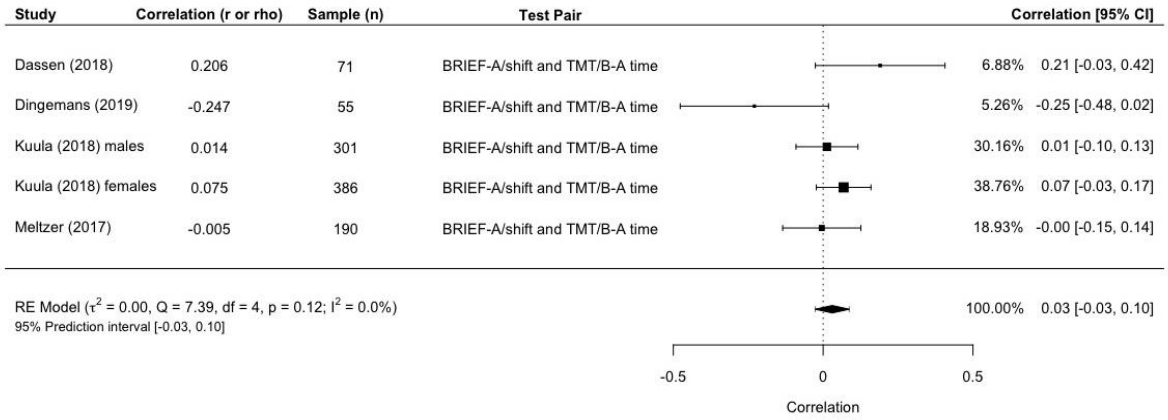


Figure S2.4. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the TMT – B-A (time).

A meta-analysis of four studies (five cohorts) indicated that the BRIEF-A shift subscale (total score) and the TMT – B-A (time) did not relate. That is, there was a non-significant effect estimate of 0.03 (95% CI = -0.03 to 0.10).

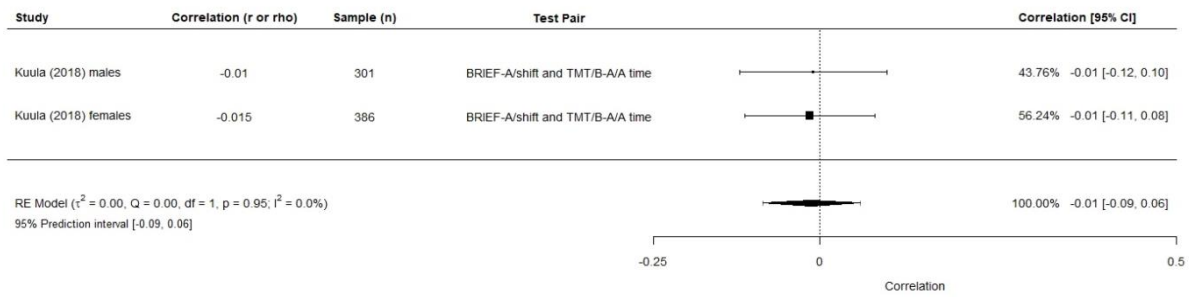


Figure S2.5. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the TMT – B-A/A (time).

A meta-analysis of one study (two cohorts) revealed that the BRIEF-A shift subscale (total score) and the TMT – B-A/A (time) did not relate. That is, the test pair did not share a relationship (*r* = -0.01, 95% CI = -0.09 to 0.06).

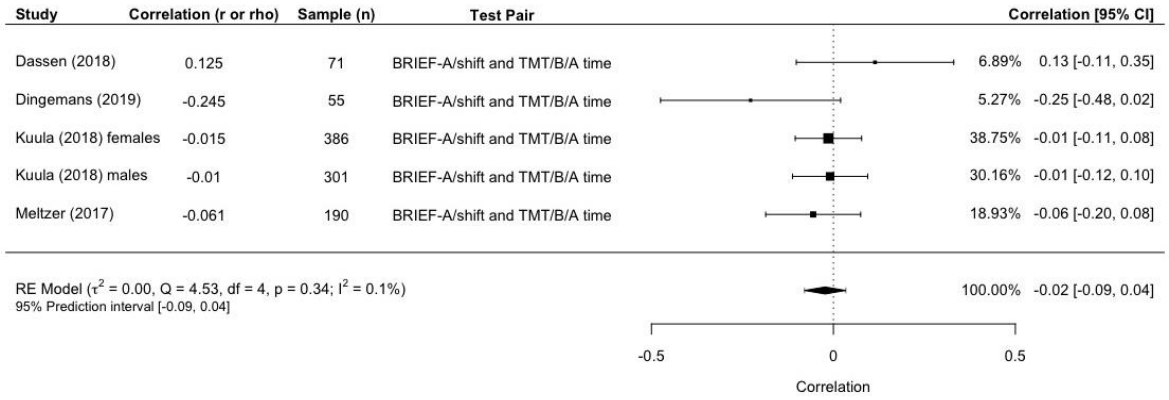


Figure S2.6. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the TMT – B/A (time).

A meta-analysis of four studies (five cohorts) indicated that the BRIEF-A shift subscale (total score) and the TMT – B/A (time) did not relate. That is, there was a non-significant effect estimate of -0.02 (95% CI = -0.09 to 0.04).

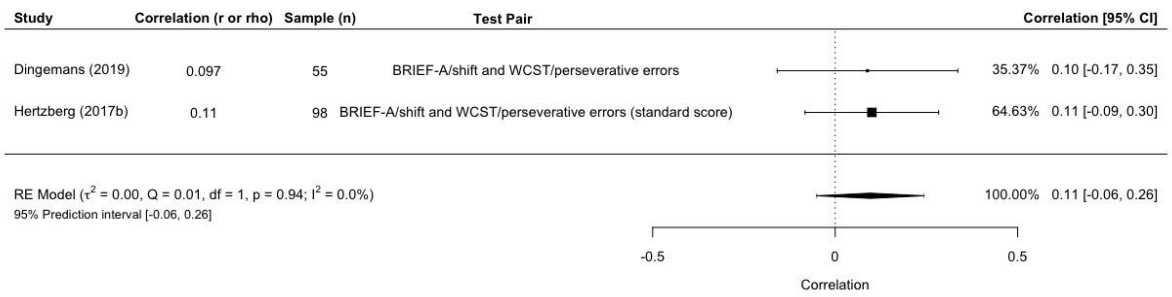


Figure S2.7. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the WCST (perseverative errors).

A meta-analysis of two studies indicated that the BRIEF-A shift subscale (total score) and the WCST (perseverative errors) did not share a relationship. That is, there was a non-significant effect estimate of 0.11 (95% CI = -0.06 to 0.26).

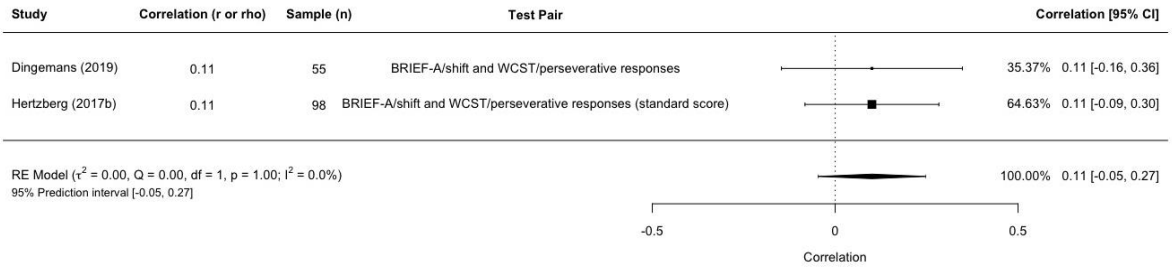


Figure S2.8. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the WCST (perseverative responses).

A meta-analysis of two studies indicated that the BRIEF-A shift subscale (total score) and the WCST (perseverative responses) did not share a relationship. That is, there was a non-significant effect estimate of 0.11 (95% CI = -0.05 to 0.27).

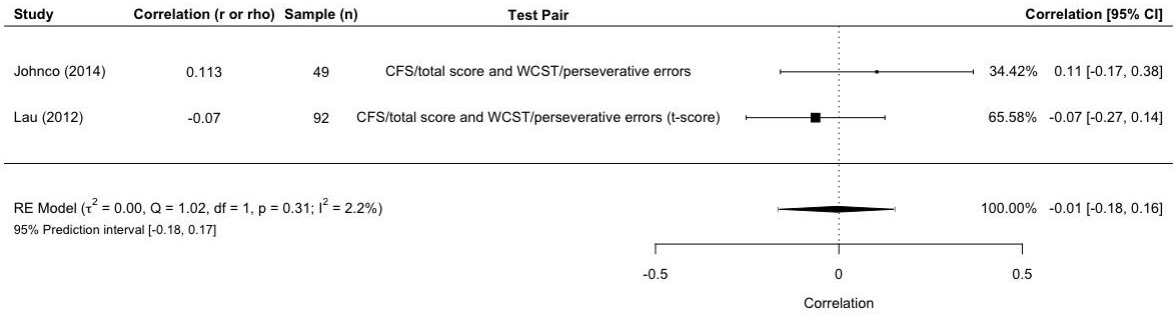


Figure S2.9. Meta-analysis of pooled cohorts that used the CFS (total score) and the WCST (perseverative errors).

A meta-analysis of two studies indicated that the CFS (total score) and the WCST (perseverative errors) did not share a relationship. That is, there was a non-significant effect estimate of -0.01 (95% CI = -0.18 to 0.16).

Appendix C: Supplementary material for Chapter Four

Howlett, C.A., Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology*, *36*(5), 347-372. https://doi.org/10.1037/neu0000796

Outline:

- **Supplementary 1 (S1):** Full details on attempts to obtain data
- **Supplementary 2 (S2):** Full details on attempts to contact authors for full-texts
- **Supplementary 3 (S3):** Meta-analyses of additional test pairs

Supplementary 1 (S1): Full details on attempts to obtain data

Reason for exclusion	Citation(s) & year
Author(s) who did not respond when contacted (*n* = 38)	Albein-Urios et al., 2018; Ballantyne, 2013; Baniqued et al., 2014; Bannon, 2017; Banz, 2014; Biederman et al., 2017; Boschloo et al., 2014; Coleman, 2010; Daly et al., 2017; Dickson et al., 2017; Dumbrava, 2017; Follmer & Stefanou, 2014; Fujino et al., 2017; García-Molina et al., 2012; García et al., 2005; Gelonch et al., 2016; Gonzalez et al., 2013; Good, 2014; Hagen et al., 2016; Healy et al., 2011; Inozemtseva et al., 2016; Kesler et al., 2013; Kesler et al., 2011; Lee, 1999; Li et al., 2013; Oswald, 2015; Pope et al., 2016; Renwick et al., 2015; Rouel et al., 2016; Sakamoto et al., 2013; Sølsnes et al., 2014; Sternheim et al., 2014; Suchy et al., 2016; Thompson, 2010; Van Noort et al., 2016; Wenner et al., 2013; Wiebe et al., 2018; Wilens et al., 2017
Contact details could not be found (*n* = 4)	Kalkut, 2010; Nordvall et al., 2017; Salman, 1999; Tamati et al., 2013
Unable to provide data (*n* = 21)	Bettis et al., 2017; Demetriou et al., 2018; Garcia-Willingham et al., 2018; Hanssen et al., 2014; Heinonen et al., 2013; Hocking et al., 2011; Lang et al., 2015; Lao-Kaim et al., 2015; Leppanen et al., 2018; Løvstad et al., 2012a; Løvstad et al., 2012b; Mangum, 2018; McAnarney et al., 2011; Mckowen et al., 2017; McKowen et al., 2018; Palm, 2004; Perkins et al., 2012; Shuster & Toplak, 2009; Svendsen et al., 2017; Tornås et al., 2016; Westwood et al., 2017
Article did not include a clinical cohort (*n* = 15)	Alfimova et al., 2017; Aytug et al., 2018; Brenner, 2017; Clayson, 2018; Cohen et al., 2017; Dawson et al., 2012; Drake, 2008; Hertzberg, 2017; Kuula et al., 2018; Lau, 2012; McCloskey, 2010; Reed et al., 2020; Saylik & Szameitat, 2018; Zmigrod et al., 2018
Author(s) were unable to determine which participants were using psychotropic medications at the time of testing (*n* = 1)	Donders et al., 2015
Duplicate data (*n* = 4)	Johnco et al., 2014; Marceau et al., 2018; McCurdy et al., 2016; Perkins et al., 2014
Participants ≥18 years did not complete self-report tests of cognitive flexibility (*n* = 1)	van Zellem et al., 2015
Author(s) were unable to confirm whether the assessment pair was completed within 1 month (*n* = 1)	Warren et al., 2013

References

Albein-Urios, N., Youssef, G. J., Kirkovski, M., & Enticott, P. G. (2018). Autism spectrum traits linked with reduced performance on self-report behavioural measures of cognitive flexibility. *Journal of Autism and Developmental Disorders*, *48*(7), 2506–2515. https://doi.org/10.1007/s10803-018-3503-3

Alfimova, M. V., Golimbet, V., Lezheiko, T., & Korovaitseva, G. (2017). A potential role of the 5-HTTLPR polymorphism in self-reported executive functioning. *Spanish Journal of Psychology*, *20*(e-13), 1–9. https://doi.org/10.1017/sjp.2017.6

Aytug, Z. G., Rua, T., Brazeal, D. V., Almaraz, J. A., & González, C. B. (2018). A socio-cultural approach to multicultural experience: Why interactions matter for creative thinking but exposures don't. *International Journal of Intercultural Relations*, *64*, 29–42. https://doi.org/10.1016/j.ijintrel.2018.03.004

Ballantyne, E. C. (2013). *The neuropsychology of mindfulness: Attentional aspects* (Publication No. 3578540) [Doctoral dissertation, Adler School of Professional Psychology]. ProQuest Dissertations and Theses Global.

Baniqued, P. L., Kranz, M. B., Voss, M. W., Lee, H., Cosman, J. D., Severson, J., & Kramer, A. F. (2014). Cognitive training with casual video games: Points to consider. *Frontiers in Psychology*, *4*(1010), 1–19. https://doi.org/10.3389/fpsyg.2013.01010

Bannon, E. E. (2017). *The efficacy of acceptance based behavior therapy versus cognitive therapy for test anxiety and working memory performance* (Publication No. 10754246) [Doctoral dissertation, Bowling Green State University]. ProQuest Dissertations and Theses Global

Banz, B. C. (2014). *An evaluation of executive functions, cognitive control and a neurocognitive profile of college binge* (Publication No. 3635581) [Doctoral dissertation, Colorado State University]. ProQuest Dissertations and Theses Global

Bettis, A. H., Coiro, M. J., England, J., Murphy, L. K., Zelkowitz, R. L., Dejardins, L., Eskridge, R., Adery, L. H., Yarboi, J., Pardo, D., & Compas, B. E. (2017). Comparison of two approaches to prevention of mental health problems in college students: Enhancing coping and executive function skills. *Journal of American College Health*, *65*(5), 313–322. https://doi.org/10.1080/07448481.2017.1312411

Biederman, J., Fried, R., Tarko, L., Surman, C., Spencer, T., Pope, A., Grossman, R., McDermott, K.,

Woodworth, K. Y., & Faraone, S. V. (2017). Memantine in the treatment of executive function deficits in adults with ADHD: A pilot-randomized double-blind controlled clinical trial. *Journal of Attention Disorders*, *21*(4), 343–352. https://doi.org/10.1177/1087054714538656

Boschloo, A., Krabbendam, L., Aben, A., De Groot, R., & Jolles, J. (2014). Sorting Test, Tower Test, and BRIEF-SR do not predict school performance of healthy adolescents in preuniversity education. *Frontiers in Psychology*, *5*(287), 1–8. https://doi.org/10.3389/fpsyg.2014.00287

Brenner, L. (2017). *The relationship between measures of cognitive flexibility and intimate partner violence perpetration in a court-mandated domestic violence treatment program* (Publication No. 10264484) [Doctoral disssertation, Rosalind Franklin University of Medicine and Science]. ProQuest Dissertations and Theses Global

Clayson, K. A. (2018). *A multimethod examination of the relevance of executive control to disgust and mental contamination among female assault survivors* (Publication No. 106173522018) [Doctoral dissertation, Baylor University]. ProQuest Dissertations and Theses Global

Cohen, J. N., Jensen, D., Stange, J. P., Neuburger, M., & Heimberg, R. G. (2017). The immediate and long-term effects of an intensive meditation retreat. *Mindfulness*, *8*(4), 1064–1077. https://doi.org/10.1007/s12671-017-0682-5

Coleman, S. L. (2010). *The relationship between cognitive inhibition and intrusive thoughts* (Publication No. 3423272) [Doctoral dissertation, State University of New York at Binghamtom]. ProQuest Dissertations and Theses Global

Daly, B. P., Hildenbrand, A. K., Turner, E., Berkowitz, S., & Tarazi, R. A. (2017). Executive functioning among college students with and without history of childhood maltreatment. *Journal of Aggression, Maltreatment and Trauma*, *26*(7), 717–735. https://doi.org/10.1080/10926771.2017.1317685

Dawson, E. L., Shear, P. K., & Strakowski, S. M. (2012). Behavior regulation and mood predict social functioning among healthy young adults. *Journal of Clinical and Experimental Neuropsychology*, *34*(3), 297–305. https://doi.org/10.1080/13803395.2011.639297

Demetriou, E. A., Song, C. Y., Park, S. H., Pepper, K. L., Naismith, S. L., Hermens, D. F., Hickie, I. B., Thomas, E. E., Norton, A., White, D., & Guastella, A. J. (2018). Autism, early psychosis, and social anxiety disorder: A transdiagnostic examination of executive function cognitive circuitry and contribution to disability. *Translational Psychiatry*, *8*(200), 1–10.

https://doi.org/10.1038/s41398-018-0193-8

Dickson, K. S., Ciesla, J. A., & Zelic, K. (2017). The role of executive functioning in adolescent rumination and depression. *Cognitive Therapy and Research*, *41*, 62–72. https://doi.org/10.1007/s10608-016-9802-0

Donders, J., Oh, Y. I., & Gable, J. (2015). Self-and informant ratings of executive functioning after mild traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *30*(6), e30–e39. https://doi.org/10.1097/HTR.0000000000000120

Drake, C. E. (2008). *Evaluating a behavioral measure of psychological flexibility* (Publication No. 3329528) [Doctoral dissertation, The University of Mississippi]. ProQuest Dissertations and Theses Global

Dumbrava, I. L. (2017). An empirical approach to wisdom processes. *Romanian Journal of Applied Psychology*, *19*(2), 50–58. https://doi.org/10.24913/rjap.19.2.04

Follmer, D. J., & Stefanou, C. R. (2014). Examining the correspondence between a direct and an indirect measure of executive functions: Implications for school-based assessment. *The School Psychologist*, *68*(3), 12–18.

Fujino, J., Tei, S., Jankowski, K. F., Kawada, R., Murai, T., & Takahashi, H. (2017). Role of spontaneous brain activity in explicit and implicit aspects of cognitive flexibility under socially conflicting situations: A resting-state fMRI study using fractional amplitude of low-frequency fluctuations. *Neuroscience*, *367*, 60–71. https://doi.org/10.1016/j.neuroscience.2017.10.025

García-Molina, A., Tormos, J. M., Bernabeu, M., Junqué, C., & Roig-Rovira, T. (2012). Do traditional executive measures tell us anything about daily-life functioning after traumatic brain injury in Spanish-speaking individuals? *Brain Injury*, *26*(6), 864–874. https://doi.org/10.3109/02699052.2012.655362

Garcia-Willingham, N. E., Roach, A. R., Kasarskis, E. J., & Segerstrom, S. C. (2018). Self-regulation and executive functioning as related to survival in Motor Neuron Disease: Preliminary findings. *Psychosomatic Medicine*, *80*(7), 665–672. https://doi.org/10.1097/PSY.0000000000000602

García, A. V., Torrecillas, F. L., Arcos, F. A. D, & García, M. P. (2005). Effects of executive impairments on maladaptive explanatory styles in substance abusers: Clinical implications. *Archives of*

Clinical Neuropsychology, *20*(1), 67–80. https://doi.org/10.1016/j.acn.2004.03.002

Gelonch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and objective measures. *Comprehensive Psychiatry*, *66*, 113–122. https://doi.org/10.1016/j.comppsych.2016.01.002

Gonzalez, C. A., Figueroa, I. J., Bellows, B. G., Rhodes, D., & Youmans, R. J. (2013). A new behavioral measure of cognitive flexibility. In D. Harris (Ed.), *International Conference on Engineering Psychology and Cognitive Ergonomics*, 297–306. Springer. https://doi.org/10.1007/978-3-642-39360-0

Good, D. (2014). Predicting real-time adaptive performance in a dynamic decision-making context. *Journal of Management and Organization*, *20*(6), 715–732. https://doi.org/10.1017/jmo.2014.54

Hagen, E., Erga, A. H., Hagen, K. P., Nesvåg, S. M., McKay, J. R., Lundervold, A. J., & Walderhaug, E. (2016). Assessment of executive function in patients with substance use disorder: A comparison of inventory- and performance-based assessment. *Journal of Substance Abuse Treatment*, *66*, 1–8. https://doi.org/10.1016/j.jsat.2016.02.010

Hanssen, K. T., Beiske, A. G., Landrø, N. I., & Hessen, E. (2014). Predictors of executive complaints and executive deficits in multiple sclerosis. *Acta Neurologica Scandinavica*, *129*(4), 234–242. https://doi.org/10.1111/ane.12177

Healy, B., Treadwell, A., & Reagan, M. (2011). Measures of RSA suppression, attentional control, and negative affect predict self-ratings of executive functions. *Journal of Psychophysiology*, *25*(4), 164–173. https://doi.org/10.1027/0269-8803/a000053

Heinonen, K., Pesonen, A. K., Lahti, J., Pyhälä, R., Strang-Karlsson, S., Hovi, P., Järvenpää, A. L., Eriksson, J. G., Andersson, S., Kajantie, E., & Raikkonen, K. (2013). Self- and parent-rated executive functioning in young adults with very low birth weight. *Pediatrics*, *131*(1), e243–e250. https://doi.org/10.1542/peds.2012-0839

Hertzberg, L. C. (2017). *The differential roles of objective neuropsychological testing and self report measures in assessing neurocognitive impairment in relation to addiction severity among alcohol abusing individuals* (Publication No. 10287393) [Doctoral dissertation, Baylor University]. ProQuest Dissertations and Theses Global

Hocking, M. C., Hobbie, W. L., Deatrick, J. A., Lucas, M. S., Szabo, M. M., Volpe, E. M., & Barakat, L. P.

(2011). Neurocognitive and family functioning and quality of life among young adult survivors of childhood brain tumors. *Clinical Neuropsychologist*, *25*(6), 942–962. https://doi.org/10.1080/13854046.2011.580284

Inozemtseva, O., Pérez-Solís, L., Matute, E., & Juárez, J. (2016). Differential improvement of executive functions during abstinence in cocaine-dependent patients: A longitudinal study. *Substance Use & Misuse*, *51*(11), 1428–1440. https://doi.org/10.1080/10826084.2016.1178293

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). The influence of cognitive flexibility on treatment outcome and cognitive restructuring skill acquisition during cognitive behavioural treatment for anxiety and depression in older adults: Results of a pilot study. *Behaviour Research and Therapy*, *57*, 55–64. https://doi.org/10.1016/j.brat.2014.04.005

Kalkut, E. (2010). *Neuropsychological assessment of executive functioning and its association with depressive symptomology* (Publication No. 3419840) [Doctoral dissertation, Loyola University Chicago]. ProQuest Dissertations and Theses Global

Kesler, S., Hadi Hosseini, S. M., Heckler, C., Janelsins, M., Palesh, O., Mustian, K., & Morrow, G. (2013). Cognitive training for improving executive function in chemotherapy-treated breast cancer survivors. *Clinical Breast Cancer*, *13*(4), 299–306. https://doi.org/10.1016/j.clbc.2013.02.004

Kesler, S. R., Kent, J. S., & O'Hara, R. (2011). Prefrontal cortex and executive function impairments in primary breast cancer. *Archives of Neurology*, *68*(11), 1447–1453. https://doi.org/10.1001/archneurol.2011.245

Kuula, L., Pesonen, A. K., Heinonen, K., Kajantie, E., Eriksson, J. G., Andersson, S., Lano, A., Lahti, J., Wolke, D., & Räikkönen, K. (2018). Naturally occurring circadian rhythm and sleep duration are related to executive functions in early adulthood. *Journal of Sleep Research*, *27*(1), 113–119. https://doi.org/10.1111/jsr.12581

Lang, K., Lloyd, S., Khondoker, M., Simic, M., Treasure, J., & Tchanturia, K. (2015). Do children and adolescents with anorexia nervosa display an inefficient cognitive processing style? *PLoS ONE*, *10*(7), 1–13. https://doi.org/10.1371/journal.pone.0131724

Lao-Kaim, N. P., Fonville, L., Giampietro, V. P., Williams, S. C. R., Simmons, A., & Tchanturia, K. (2015). Aberrant function of learning and cognitive control networks underlie inefficient cognitive flexibility in anorexia nervosa: A cross-sectional fMRI study. *PLoS ONE*, *10*(5), 1–16.

https://doi.org/10.1371/journal.pone.0124027

Lau, K. (2012). *The role of cognitive flexibility in acculturation and bicultural stress among Chinese American adults* (Publication No. 3517364) [Doctoral dissertation, Alliant International University]. ProQuest Dissertations and Theses Global

Lee, J. (1999). *Social intelligence: Flexibility and complexity are distinct from creativity* (Publication No. 9924645) [Doctoral dissertation, University of Notre Dame]. ProQuest Dissertations and Theses Global

Leppanen, J., Adamson, J., & Tchanturia, K. (2018). Impact of cognitive remediation therapy on neurocognitive processing in anorexia nervosa. *Frontiers in Psychiatry*, *9*(96), 1–5. https://doi.org/10.3389/fpsyt.2018.00096

Li, Y., Dong, F., Cao, F., Cui, N., Li, J., & Long, Z. (2013). Poly-victimization and executive functions in junior college students. *Scandinavian Journal of Psychology*, *54*(6), 485–492. https://doi.org/10.1111/sjop.12083

Løvstad, M., Funderud, I., Endestad, T., Due-Tonnessen, P., Meling, T. R., Lindgren, M., Knight, R. T., & Solbakk, A. K. (2012a). Executive functions after orbital or lateral prefrontal lesions: Neuropsychological profiles and self-reported executive functions in everyday living. *Brain Injury*, *26*(13–14), 1586–1598. https://doi.org/10.3109/02699052.2012.698787

Løvstad, M., Funderud, I., Meling, T., Krämer, U. M., Voytek, B., Due-Tønnessen, P., Endestad, T., Lindgren, M., Knight, R. T., & Solbakk, A. K. (2012b). Anterior cingulate cortex and cognitive control: Neuropsychological and electrophysiological findings in two patients with lesions to dorsomedial prefrontal cortex. *Brain and Cognition*, *80*(2), 237–249. https://doi.org/10.1016/j.bandc.2012.07.008

Mangum, R. W. (2018). *Self-understanding of executive functions in individuals with agenesis of the corpus callosum* (Publication No. 10792935) [Doctoral dissertation, Fuller Theological Seminary, School of Psychology]. ProQuest Dissertations and Theses Global

Marceau, E. M., Kelly, P. J., & Solowij, N. (2018). The relationship between executive functions and emotion regulation in females attending therapeutic community treatment for substance use disorder. *Drug and Alcohol Dependence*, *182*, 58–66. https://doi.org/10.1016/j.drugalcdep.2017.10.008

McAnarney, E. R., Zarcone, J., Singh, P., Michels, J., Welsh, S., Litteer, T., Wang, H., & Klein, J. D.

(2011). Restrictive anorexia nervosa and set-shifting in adolescents: A biobehavioral interface. *Journal of Adolescent Health*, *49*(1), 99–101. https://doi.org/10.1016/j.jadohealth.2010.11.259

McCloskey, L. (2010). Response frequency and Exner's active-passive ratio as Rorschach indicators of flexibility. *Perceptual and Motor Skills*, *111*(2), 559–575. https://doi.org/10.2466/02.03.04.09.PMS.111.5.559-575

McCurdy, M. D., Turner, E. M., Barakat, L. P., Hobbie, W. L., Deatrick, J. A., Paltin, I., Fisher, M. J., & Hocking, M. C. (2016). Discrepancies among measures of executive functioning in a subsample of young adult survivors of childhood brain tumor: Associations with treatment intensity. *Journal of the International Neuropsychological Society*, *22*(9), 900–910. https://doi.org/10.1017/S1355617716000771

McKowen, J., Carrellas, N., Zulauf, C., Ward, E. N., Fried, R., & Wilens, T. (2017). Factors associated with attrition in substance using patients enrolled in an intensive outpatient program. *American Journal on Addictions*, *26*(8), 780–787. https://doi.org/10.1111/ajad.12619

McKowen, J. W., Isenberg, B. M., Carrellas, N. W., Zulauf, C. A., Ward, N. E., Fried, R. S., & Wilens, T. E. (2018). Neuropsychological changes in patients with substance use disorder after completion of a one month intensive outpatient treatment program. *American Journal on Addictions*, *27*(8), 632–638. https://doi.org/10.1111/ajad.12824

Nordvall, O., Jonsson, B., & Neely, A. S. (2017). Self-reported and performance-based measures of executive functions in interned youth. *Psychology, Crime and Law*, *23*(3), 240–253. https://doi.org/10.1080/1068316X.2016.1239725

Oswald, A. (2015). *Effects of repeated exposure to stressors on cognitive flexibility as a function of self-type and self-complexity: A multi-session experimental test* [Unpublished Doctoral dissertation]. University of Newcastle.

Palm, K. M. (2004). *An examination of mindfulness: Assessment and relationship to PTSD* (Publication No. 3161689) [Doctoral dissertation, University of Nevada, Reno]. ProQuest Dissertations and Theses Global

Perkins, S. C., Cortina, K. S., Smith-Darden, J. P., & Graham-Bermann, S. A. (2012). The mediating role of self-regulation between intrafamilial violence and mental health adjustment in incarcerated male adolescents. *Journal of Interpersonal Violence*, *27*(7), 1199–1224.

https://doi.org/10.1177/0886260511425240

Perkins, S. C., Smith-Darden, J., Ametrano, R. M., & Graham-Bermann, S. (2014). Typologies of violence exposure and cognitive processing in incarcerated male adolescents. *Journal of Family Violence*, *29*(4), 439–451. https://doi.org/10.1007/s10896-014-9600-y

Pope, C. N., Ross, L. A., & Stavrinos, D. (2016). Association between executive function and problematic adolescent driving. *Journal of Developmental and Behavioral Pediatrics*, *37*(9), 702–711. https://doi.org/10.1097/DBP.0000000000000353

Reed, R. G., Combs, H. L., & Segerstrom, S. C. (2020). The structure of self-regulation and its psychological and physical health correlates in older adults. *Collabra: Psychology*, *6*(1), 23. https://doi.org/10.1525/collabra.297

Renwick, B., Musiat, P., Lose, A., Dejong, H., Broadbent, H., Kenyon, M., Loomes, R., Watson, C., Ghelani, S., Serpell, L., Richards, L., Johnson-Sabine, E., Boughton, N., Treasure, J., & Schmidt, U. (2015). Neuro- and social-cognitive clustering highlights distinct profiles in adults with anorexia nervosa. *International Journal of Eating Disorders*, *48*(1), 26–34. https://doi.org/10.1002/eat.22366

Rouel, M., Raman, J., Hay, P., & Smith, E. (2016). Validation of the Behaviour Rating Inventory of Executive Function-Adult Version (BRIEF-A) in the obese with and without binge eating disorder. *Eating Behaviors*, *23*, 58–65. https://doi.org/10.1016/j.eatbeh.2016.07.010

Sakamoto, M., Woods, S. P., Kolessar, M., Kriz, D., Anderson, J. R., Olavarria, H., Sasaki, A. W., Chang, M., Flora, K. D., Loftis, J. M., & Huckans, M. (2013). Protective effects of higher cognitive reserve for neuropsychological and daily functioning among individuals infected with hepatitis C. *Journal of NeuroVirology*, *19*(5), 442–451. https://doi.org/10.1007/s13365-013-0196-4

Salmon, D. (1999). *Attentional control and the deconstruction of the schemas underlying the perception of pain* (Publication No. 9934176) [Doctoral dissertation, Long Island University]. ProQuest Dissertations and Theses Global

Saylik, R., & Szameitat, A. J. (2018). The association between negative attributional style and working memory performance. *The Open Psychology Journal*, *11*, 131–141. https://doi.org/10.2174/1874350101811010131

Shuster, J., & Toplak, M. E. (2009). Executive and motivational inhibition: Associations with self-

report measures related to inhibition. *Consciousness and Cognition*, *18*(2), 471–480. https://doi.org/10.1016/j.concog.2009.01.004

Sølsnes, A. E., Skranes, J., Brubakk, A. M., & Lohaugen, G. C. C. (2014). Executive functions in very-low-birth-weight young adults: A comparison between self-report and neuropsychological test results. *Journal of the International Neuropsychological Society*, *20*(5), 506–515. https://doi.org/10.1017/S1355617714000332

Sternheim, L., van der Burgh, M., Berkhout, L. J., Dekker, M. R., & Ruiter, C. (2014). Poor cognitive flexibility, and the experience thereof, in a subclinical sample of female students with obsessive-compulsive symptoms. *Scandinavian Journal of Psychology*, *55*(6), 573–577. https://doi.org/10.1111/sjop.12163

Suchy, Y., Turner, S. L., Queen, T. L., Durracio, K., Wiebe, D. J., Butner, J., Franchow, E. I., White, P. C., Murray, M. A., Swinyard, M., & Berg, C. A. (2016). The relation of questionnaire and performance-based measures of executive functioning with type 1 diabetes outcomes among late adolescents. *Health Psychology*, *35*(7), 661– 669. https://doi.org/10.1037/hea0000326

Svendsen, T. S., Erga, A. H., Hagen, E., McKay, J. R., Njå, A. L. M., Årstad, J., & Nesvåg, S. (2017). How to maintain high retention rates in long-term research on addiction: A case report. *Journal of Social Work Practice in the Addictions*, *17*(4), 374–387. https://doi.org/10.1080/1533256X.2017.1361831

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: A first report. *Journal of the American Academy of Audiology*, *24*(7), 616–634. https://doi.org/10.3766/jaaa.24.7.10

Thompson, R. (2010). *Understanding bulimia nervosa from a neuropsychological perspective: Impulsivity and binge-purge behavior in adolescent and young adult women* (Publication No. 3372002) [Doctoral dissertation, Temple University]. ProQuest Dissertations and Theses Global

Tornås, S., Lovstad, M., Solbakk, A. K., Evans, J., Endestad, T., Hol, P. K., Schanke, A. K., & Stubberud, J. (2016). Rehabilitation of executive functions in patients with chronic acquired brain injury with goal management training, external cuing, and emotional regulation: A randomized controlled trial. *Journal of the International Neuropsychological Society*, *22*(4), 436–452. https://doi.org/10.1017/S1355617715001344

Van Noort, B. M., Kraus, M. K. A., Pfeiffer, E., Lehmkuhl, U., & Kappel, V. (2016). Neuropsychological and behavioural short-term effects of cognitive remediation therapy in adolescent anorexia nervosa: A pilot study. *European Eating Disorders Review*, *24*(1), 69–74. https://doi.org/10.1002/erv.2383

van Zellem, L., Buysse, C., Madderom, M., Legerstee, J. S., Aarsen, F., Tibboel, D., & Utens, E. M. (2015). Long-term neuropsychological outcomes in children and adolescents after cardiac arrest. *Intensive Care Medicine*, *41*(6), 1057–1066. https://doi.org/10.1007/s00134-015-3789-y

Warren, S. L., Crocker, L. D., Spielberg, J. M., Engels, A. S., Banich, M. T., Sutton, B. P., Miller, G. A., & Heller, W. (2013). Cortical organization of inhibition-related functions and modulation by psychopathology. *Frontiers in Human Neuroscience*, *7*(271), 1–13. https://doi.org/10.3389/fnhum.2013.00271

Wenner, C. J., Bianchi, J., Figueredo, A. J., Rushton, J. P., & Jacobs, W. J. (2013). Life history theory and social deviance: The mediating role of executive function. *Intelligence*, *41*(2), 102–113. https://doi.org/10.1016/j.intell.2012.11.004

Westwood, H., Mandy, W., & Tchanturia, K. (2017). The association between symptoms of autism and neuropsychological performance in females with anorexia nervosa. *Psychiatry Research*, *258*, 531–537. https://doi.org/10.1016/j.psychres.2017.09.005

Wiebe, D. J., Baker, A. C., Suchy, Y., Stump, T. K., & Berg, C. A. (2018). Individual differences and day-to-day fluctuations in goal planning and type 1 diabetes management. *Health Psychology*, *37*(7), 638–646. https://doi.org/10.1037/hea0000624

Wilens, T. E., Carrellas, N. W., Martelon, M. K., Yule, A. M., Fried, R., Anselmo, R., & McCabe, S. E. (2017). Neuropsychological functioning in college students who misuse prescription stimulants. *American Journal on Addictions*, *26*(4), 379–387. https://doi.org/10.1111/ajad.12551

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2018). Cognitive underpinnings of nationalistic ideology in the context of Brexit. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), E4532–E4540. https://doi.org/10.1073/pnas.1708960115

Supplementary 2 (S2): Full details on attempts to contact authors for full-texts

Attempts were made to contact 112 authors to obtain full-texts. Of these 112 authors, we attempted to contact 108 by either email or ResearchGate (ResearchGate GmbH, Berlin, Germany), and four were contacted via phone. We allowed two weeks for these authors to provide either a published paper, unpublished manuscript and/or the raw or aggregate data. Of these 112 records, 70 were irretrievable because either the author(s) contact details could not be found, they did not reply, or they were unable to provide the necessary data. For the remaining 42 records, the authors were able to provide sufficient information to determine study eligibility, but none were eligible for inclusion. Any non-English publications were translated using Google Translate. Almost all records were excluded after attempting to contact authors to obtain full-texts (see Box 1 for full-text exclusions (with reasons) in Figure 1); however, three records were not excluded until after clarifying particular aspects of the eligibility criteria and prior to requesting data (see Box 2 for full-text exclusions (with reasons) in Figure 1). Thus, for three records the authors were contacted twice.

Supplementary 3 (S3): Meta-analyses of additional test pairs

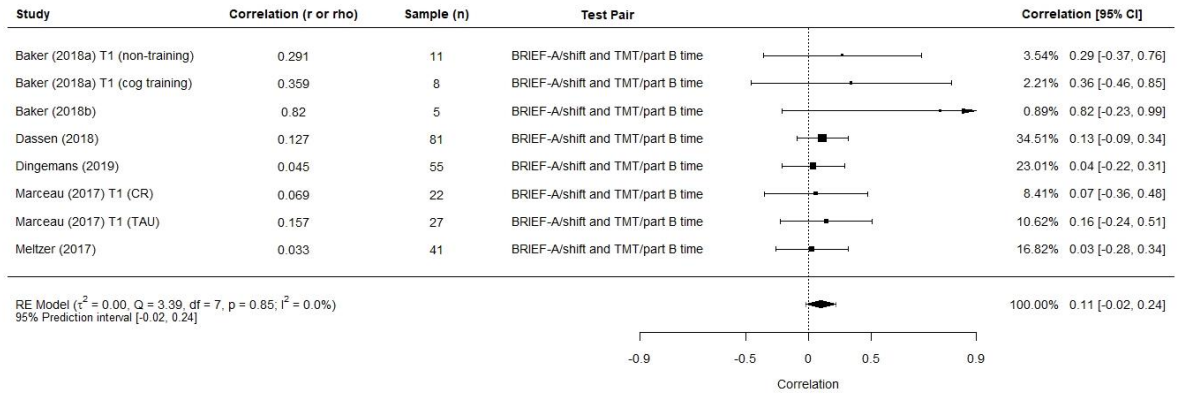


Fig. S3.1. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the TMT – part B (time).

A meta-analysis of 6 studies (8 comparisons) revealed that the BRIEF-A shift subscale (total score) and the TMT – part B (time) did not relate (*r* = 0.11, 95% CI = -0.02 to 0.24).

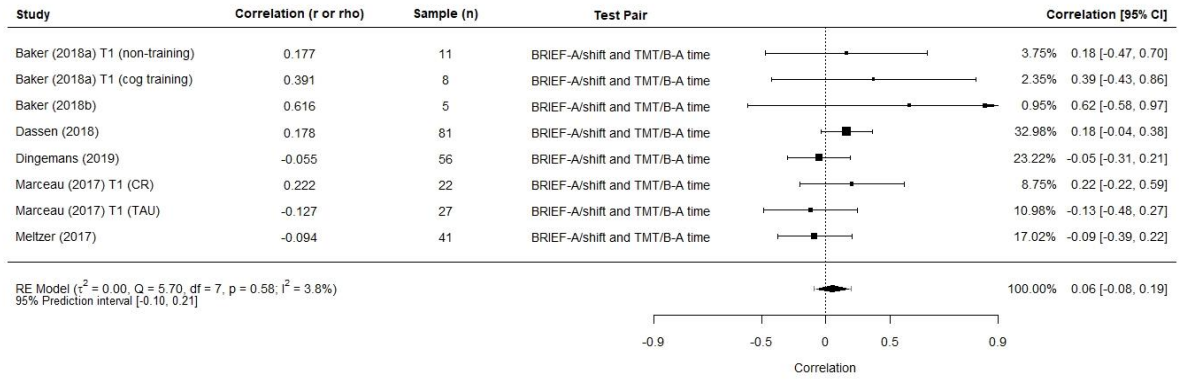


Fig. S3.2. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the TMT – B-A (time).

A meta-analysis of 6 studies (8 comparisons) revealed that the BRIEF-A shift subscale (total score) and the TMT – B-A (time) did not correlate. That is, the test pair did not share a relationship (*r* = 0.06, 95% CI = -0.08 to 0.19).

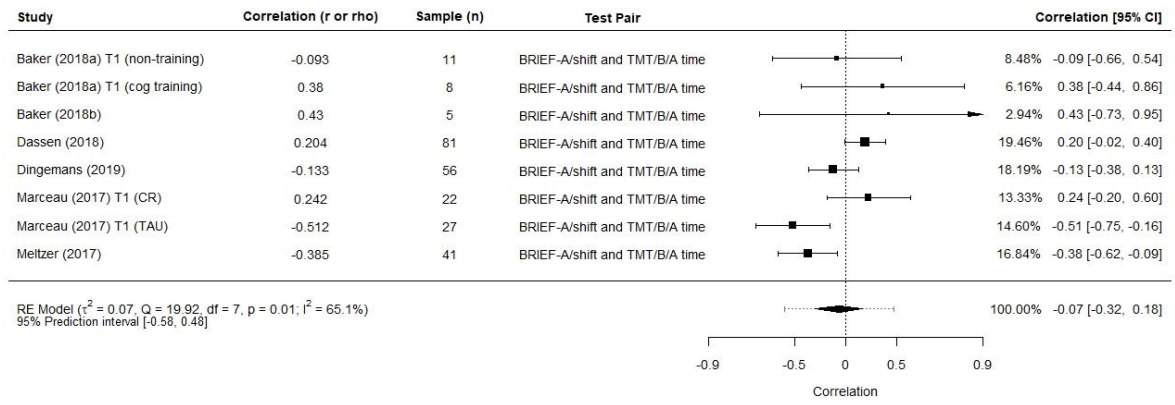


Fig. S3.3. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the TMT – B/A (time).

A meta-analysis of 6 studies (8 comparisons) revealed that the BRIEF-A shift subscale (total score) and the TMT – B/A (time) did not relate. That is, the test pair did not share a relationship (*r* = -0.07, 95% CI = -0.32 to 0.18). Statistical heterogeneity was detected for this test pair (Q = 19.92, *p* = 0.01, I^2 = 65.1%).

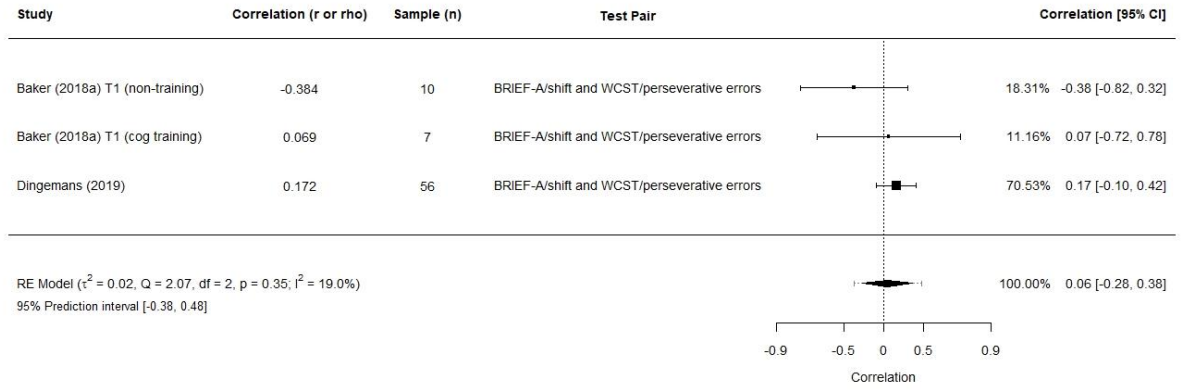


Fig. S3.4. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the WCST (perseverative errors).

A meta-analysis of 2 studies (3 comparisons) indicated that the BRIEF-A shift subscale (total score) and the WCST (perseverative errors) did not share a relationship. That is, there was a non-significant effect estimate of 0.06 (95% CI = -0.28 to 0.38).

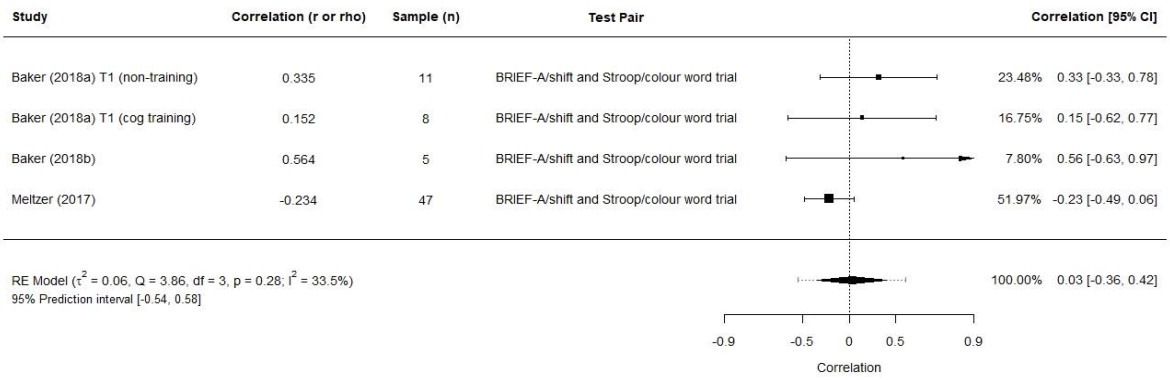


Fig. S3.5. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the Stroop (Colour-Word trial).

A meta-analysis of 3 studies (4 comparisons) indicated that the BRIEF-A shift subscale (total score) and the Stroop (Colour-Word trial) did not relate. That is, there was a non-significant effect estimate of 0.03 (95% CI = -0.36 to 0.42).

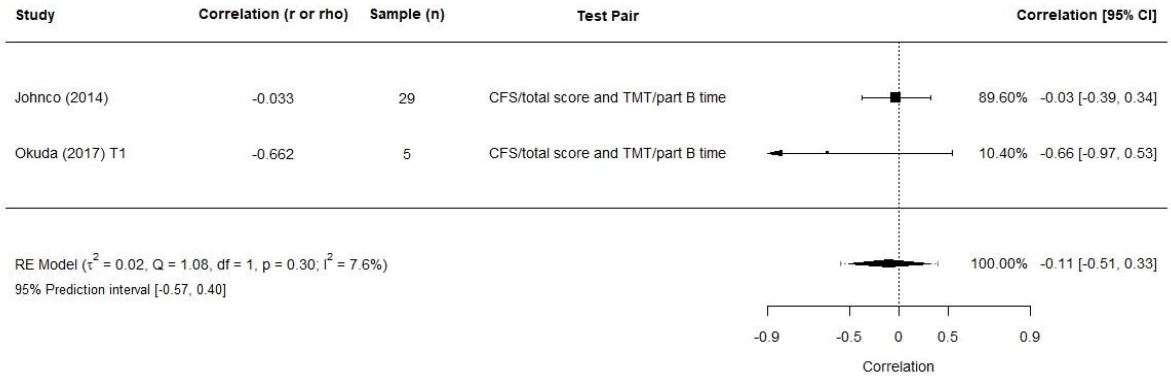


Fig. S3.6. Meta-analysis of pooled cohorts that used the CFS (total score) and the TMT – part B (time).

A meta-analysis of 2 studies (2 comparisons) indicated that the CFS (total score) and the TMT – part B (time) do not share a relationship. That is, there was a non-significant effect estimate of -0.11 (95% CI = -0.51 to 0.33).

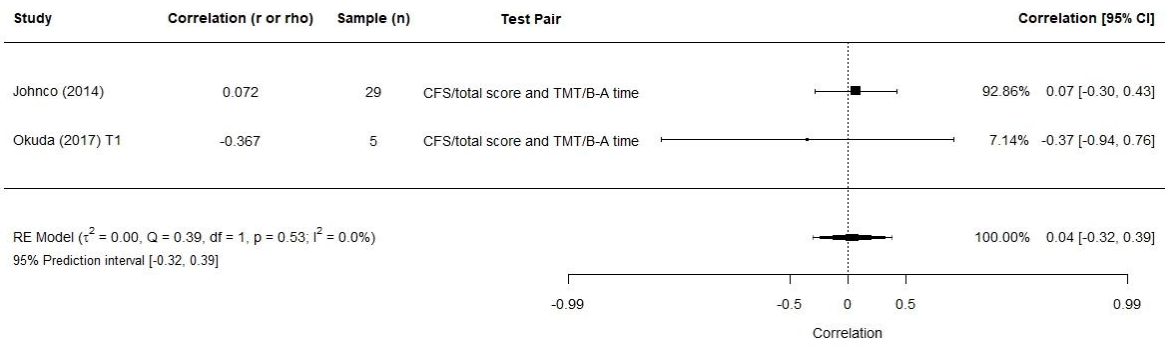


Fig. S3.7. Meta-analysis of pooled cohorts that used the CFS (total score) and the TMT –B-A (time).

A meta-analysis of 2 studies (2 comparisons) indicated that the CFS (total score) and the TMT – B-A (time) did not relate. That is, there was a non-significant effect estimate of 0.04 (95% CI = -0.32 to 0.39).

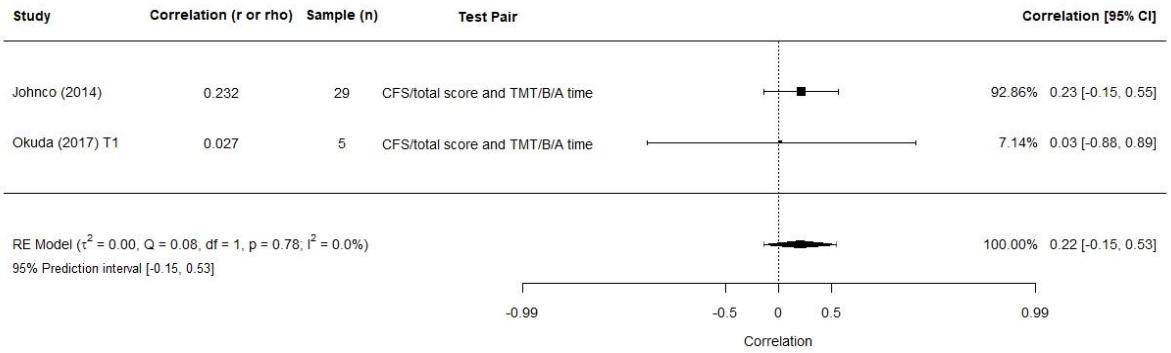


Fig. S3.8. Meta-analysis of pooled cohorts that used the CFS (total score) and the TMT – B/A (time).

A meta-analysis of 2 studies (2 comparisons) indicated that the CFS (total score) and the TMT – B/A (time) did not correlate. That is, there was a non-significant effect estimate of 0.22 (95% CI = -0.15 to 0.53).

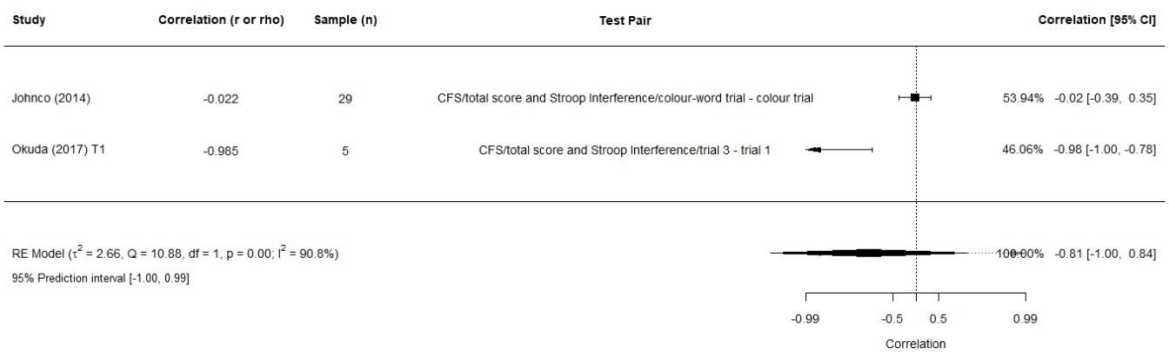


Fig. S3.9. Meta-analysis of pooled cohorts that used the CFS (total score) and the Stroop (interference score).

A meta-analysis of 2 studies (2 comparisons) indicated that the CFS (total score) and the Stroop (interference score) did not share a relationship. That is, there was a non-significant effect estimate of -0.81 (95% CI = -1.00 to 0.84). Statistical heterogeneity was detected for this test pair (Q = 10.88, p = 0.00, I^2 = 90.8%).

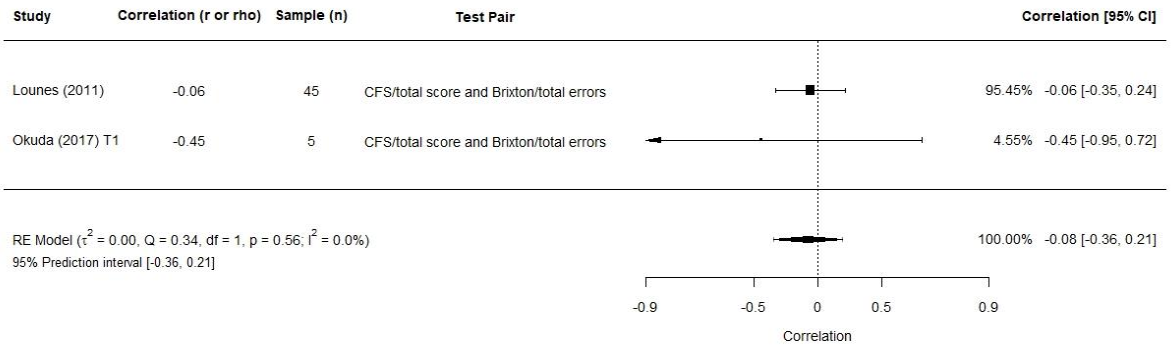


Fig. S3.10. Meta-analysis of pooled cohorts that used the CFS (total score) and the Brixton Spatial Anticipation Task (total errors).

A meta-analysis of 2 studies (2 comparisons) indicated that the CFS (total score) and the Brixton Spatial Anticipation Task (total errors) did not correlate. That is, there was a non-significant effect estimate of -0.08 (95% CI = -0.36 to 0.21).

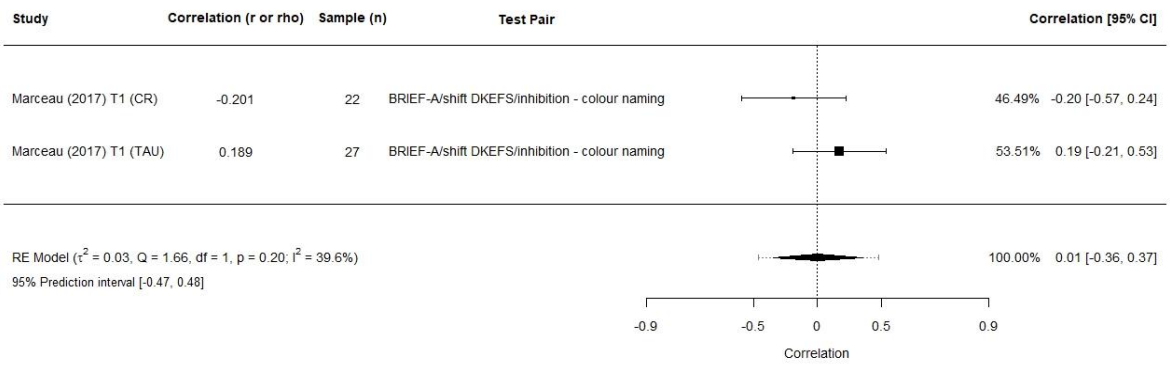


Fig. S3.11. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the D-KEFS (inhibition - colour naming score).

A meta-analysis of 1 study (2 comparisons) indicated that the BRIEF-A shift subscale (total score) and the D-KEFS (inhibition - colour naming score) did not relate. That is, there was a non-significant effect estimate of 0.01 (95% CI = -0.36 to 0.37).

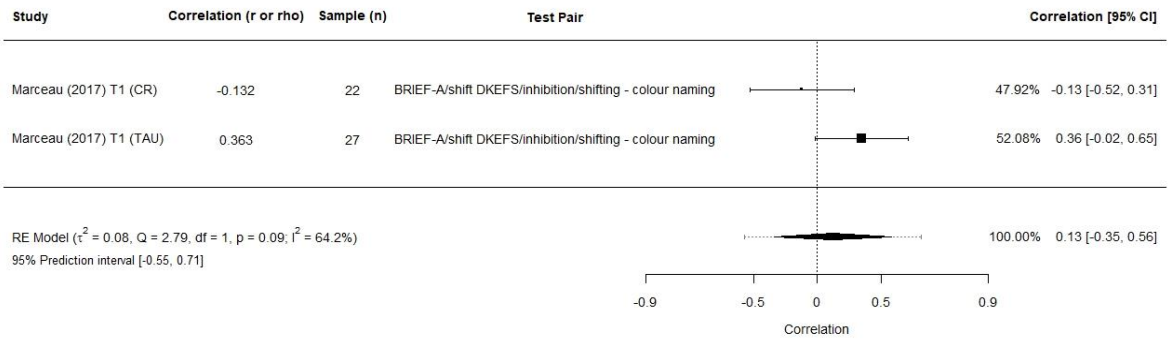


Fig. S3.12. Meta-analysis of pooled cohorts that used the BRIEF-A shift subscale (total score) and the D-KEFS (inhibition/switching - colour naming score).

A meta-analysis of 1 study (2 comparisons) indicated that the BRIEF-A shift subscale (total score) and the D-KEFS (inhibition/switching - colour naming score) did not relate. That is, there was a non-significant effect estimate of 0.13 (95% CI = -0.35 to 0.56). Statistical heterogeneity was detected for this test pair (Q = 2.79, p = 0.09, I^2 = 64.2%).

Appendix D: Supplementary material for Chapter Six

Howlett, C.A., Stanford, T., Berryman, C., Karran, E.L., Bellan, V., Coussens, S., Miles, S., & Moseley, G.L. Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study.

Outline:

- **Supplementary 1 (S1):** Pre-registered protocol
- **Supplementary 2 (S2):** STROBE Statement Checklist
- **Supplementary 3 (S3):** CHERRIES Checklist
- **Supplementary 4 (S4):** Recruitment flyer
- **Supplementary 5 (S5):** Advertisements for social media platforms and websites
- **Supplementary 6 (S6):** Post for UniSA's Research Volunteers website
- **Supplementary 7 (S7):** Recruitment email for the Research Participation Sign-up System
- **Supplementary 8 (S8):** Information sheet and consent form
- **Supplementary 9 (S9):** Debriefing form/completion page
- **Supplementary 10 (S10):** Online survey
- **Supplementary 11 (S11):** Participant demographics (continued)
- **Supplementary 12 (S12):** Diagnostic plots for multiple linear regression models
- **Supplementary 13 (S13):** Graphical depiction of the estimates for the final model parameters
- **Supplementary 14 (S14):** Side-by-side graphical depiction of the estimates for the final model parameters
- **Supplementary 15 (S15):** Boxplots depicting outliers for multiple linear regression models
- **Supplementary 16 (S16):** Diagnostic plots for correlation analysis
- **Supplementary 17 (S17):** Boxplots depicting outliers for correlation analysis
- **Supplementary 18 (S18):** Diagnostic plots for logistic regression model
- **Supplementary 19 (S19):** Statistical analysis plan
- **Supplementary 20 (S20):** Deviations from statistical analysis plan

Supplementary 1 (S1): Pre-registered protocol

Registered on Open Science Framework on 10th August 2021: www.doi.org/10.17605/OSF.IO/9Q8UT

Study Protocol

Cognitive flexibility and pain: A protocol for an online, cross-sectional, observational study.

Caitlin A. Howlett, Ty Stanford, Scott Coussens, Carolyn Berryman, Valeria Bellan, Emma L. Karran, Stephanie Miles & G. Lorimer Moseley

---

Background

Pain is a complex phenomenon.[1] Addressing the various biological, psychological and contextual factors that have the potential to contribute to the experience of pain, and as such treating persistent pain from a biopsychosocial perspective, is currently viewed as the most efficacious approach.[2] Given that persistent pain affects approximately 20% of adults globally,[3] understanding and uncovering potential contributors that make some people more susceptible to the development, maintenance and recovery of persistent pain is imperative. One possible therapeutic target is *cognitive flexibility* – a latent construct that has been broadly described as the ability to modify cognitive and behavioural strategies in accordance to changing task demands or unexpected conditions in the environment.[4,5] Cognitive flexibility has typically been assessed using self-report tests such as the Cognitive Flexibility Scale[6,7] and the Cognitive Flexibility Inventory[8], as well as neurocognitive tasks including the Trail Making Test[9,10] and the Wisconsin Card Sorting Test[11,12].

Various studies have demonstrated that people with anorexia nervosa[13,14], major depressive disorder[15] and comorbid anxiety and depression[16], have poorer neurocognitive and self-reported cognitive flexibility than healthy controls. Additionally, poor neurocognitive cognitive flexibility predicts pain chronicity in people undergoing surgery for knee osteoarthritis or breast cancer,[17]. Despite some disparities in results between different test outcomes, previous systematic reviews have consistently reported small to moderate effects between people with and without persistent pain on neurocognitive tasks of cognitive flexibility.[18–21] Despite the fact that there is mounting evidence that people with persistent pain experience problems on neurocognitive tasks, only one study[22] has investigated whether people with persistent pain – that is, fibromyalgia – are worse than healthy controls on the two assessment approaches, specifically neurocognitive tasks and self-reported tests that purport to assess cognitive flexibility. However, the study found no group differences after controlling for depression and anxiety.

The primary aim of this online study is to determine how people with persistent pain perform on neurocognitive and self-reported cognitive flexibility assessments, relative to healthy controls.

Primary research questions:

1. Do people with persistent pain perform worse than healthy controls on a neurocognitive assessment of cognitive flexibility?
2. Is self-reported cognitive flexibility lower in people with persistent pain than it is in healthy controls?

Although there is evidence that people with conditions such as anorexia nervosa perform worse, on average, than healthy controls on both self-report and neurocognitive tasks of cognitive flexibility[13,14], two comprehensive meta-analyses found no relationship between self-report and neuropsychological tests of cognitive flexibility in non-clinical[23] and clinical[24] cohorts. Similarly, another study found no relationship between self-report and neuropsychological tests of cognitive flexibility in people with and without

fibromyalgia[22]. One potential explanation for the lack of relationship observed between self-report and neurocognitive tests is that there may be particular sub-groups of people who are differentially inflexible on these two types of assessments, which remains to be elucidated in people with persistent pain.

Secondary research questions:

- Is there a relationship between a self-report and neurocognitive test of cognitive flexibility in people with persistent pain? (If so, what is the strength and direction of the association?)
- Is there a relationship between a self-report and neurocognitive test of cognitive flexibility in healthy controls? (If so, what is the strength and direction of the association?)

Exploratory research question:

- Is the probability of having persistent pain associated with a person's scores on a self-report and neurocognitive assessment of cognitive flexibility?

Methods

Design

This study will implement an online, cross-sectional, observational design.

Participants

Recruitment

Participants will be recruited via email, advertising on social media platforms (e.g., Facebook, Twitter) and recruitment flyers displayed on noticeboards at the University of South Australia's City East and Magill Campuses. To be involved in the study, participants will voluntarily follow a link to the online survey.

Study inclusion criteria

Individuals will be eligible for inclusion if they (1) live with or without persistent pain (i.e., pain experienced most days for 3 months or longer),[25] (2) are aged ≥ 18 years, (3) can read and understand English, (4) have access to, and competency with, a computer/laptop or smartphone device with internet access and (5) have normal or corrected to normal (e.g., glasses or contact lenses) vision.

Study exclusion criteria

Individuals will be excluded if they are colour blind.

Independent variables

- Group: persistent pain vs healthy controls

Covariates

- Test administration order: order 1 (CFI first, WCST last) vs order 2 (WCST first, CFI last)
- Modality of survey completion: computer/laptop vs smartphone
- Age
- Education level

- Medication use
- Gender
- **The Hospital Anxiety and Depression Scale (HADS) (Zigmond & Snaith, 1983)**
    - The Hospital Anxiety and Depression Scale (HADS)[26] is a self-report questionnaire designed to capture anxiety and depression in general clinical populations. The scale contains 14 items (7 items for the depression score and 7 items for the anxiety score), each rated on a 4-point Likert scale, with higher scores reflecting more severe symptoms.
- **Immune Status Questionnaire (ISQ) (Wilod Versprille et al., 2019)**
    - The Immune Status Questionnaire (ISQ)[27] is a self-report questionnaire designed to assess one's immune status over the past 12 months. The ISQ contains 7 items that are scored on a 5-point Likert scale from 0 (never) to 4 (almost always), with lower scores indicating poor perceived immune status. It is also recommended that current immune and health status is assessed in conjunction with the ISQ.[27] Data from the ISQ will be collected as part of a larger study.

Dependent variables

Primary outcome(s)

- **Cognitive Flexibility Inventory (Dennis & Vander Wal, 2010)**
    - The Cognitive Flexibility Inventory (CFI) is a 20-item self-report questionnaire originally developed for the purposes of assessing the ability to think more adaptively, rather than maladaptively, when faced with stressful life events[8]. Questionnaire items include, for example, "I have a hard time making decisions when faced with difficult situations". The CFI is comprised of two subscales (1) *Alternatives* (13 items) and (2) *Control* (7 items), and responses range on a 7-point Likert scale where 1 = *strongly disagree* and 7 = *strongly agree.* Higher scores are intended to represent higher cognitive flexibility; lower scores are intended to represent lower cognitive flexibility. The total score (calculated as a proportion) will be used as the outcome variable and as an indicator of self-evaluated cognitive flexibility. We selected the CFI because of its clinical relevance and utility.[8]
- **Wisconsin Card Sorting Test (Heaton et al. 1993)**
    - The Wisconsin Card Sorting Test (WCST) is a popular neurocognitive task for assessing cognitive flexibility[16,28,29]. A computerised version of the WCST that is scored in accordance to the standardised methods outlined by Heaton et al.[12] will be used. One-hundred and twenty-eight response cards are matched to four stimulus cards that vary on three dimensions – colour, form and number. Through trial and error, participants must establish the correct sorting rule, and participants receive feedback (i.e., 'correct' or 'incorrect') after each response. After 10 correct responses in a row, the correct sorting rule changes and the participant is required to change their sorting strategy (i.e., sort the cards by a different dimension).[12] The percentage of perseverative responses (calculated as a proportion) will be used as the outcome variable and as an indicator of neurocognitive cognitive flexibility[30], where larger scores indicate poorer performance[12]. We selected the WCST due to its wide acceptance as a neurocognitive task of cognitive flexibility, particularly perseverative responses as an outcome variable.[8,30]

Data Collection and Storage

Participants will access the online survey via the Inquisit Web software hosted online by Millisecond (Inquisit 6, 2020) (https://www.millisecond.com/)[31] using a personal device (i.e., computer/laptop or smartphone). All survey responses and research data will be collected using the Inquisit software and managed in accordance with the Australian Code for Conduct of Research. Each participant will receive a re-identifiable ID code and all electronic data will be password protected and securely stored on UniSA's repository to ensure the anonymity of participants.

Ethical Considerations

An online ethics application will be submitted, and approval will be sought from the University of South Australia's Human Research Ethics Committee (HREC). Involvement does not pose any identifiable risks to study participants. Participants will be provided with an information sheet that outlines their involvement in the study, and will confirm their consent by commencing the online survey. Participants will be informed that they have the option to withdraw from the study at any time without penalty.

Procedure

The purpose of the study will be outlined upon accessing the online survey via a link provided in the recruitment email/advertisement. Participants will consent to their involvement by voluntarily commencing the survey. Participants will complete some basic screening questions to ensure that consenting participants are eligible for inclusion. Such questions will include: (1) are you at least 18 years of age? (2) are you colour-blind? (3) can you read and understand English? Eligible participants will then proceed to complete a 15-item questionnaire that specifically collects information regarding the participants demographics, health and pain. Participants who identify as having persistent pain will be asked an additional five questions about their pain. All participants will then complete the CFI, the HADS, the ISQ and WCST. Tests will be counterbalanced such that some participants will have the tests/questionnaires administered in the following order: CFI, HADS, ISQ and WCST, while other participants will have the tests/questionnaires administered in the reversed order. All questionnaires and neurocognitive tests will be administered via a web version of Inquisit.[31]

Statistical analysis

Data will be statistically analysed using R.[32] Unless otherwise indicated, the level of significance will be set at $p<0.05$. Any data collected prior to participant withdrawal (i.e., non-completion of primary outcomes as per HREC (protocol no. 202439)) will not be analysed. Where duplicate participants (i.e., same user ID) are present in the dataset, the first entry will be retained for analysis and any subsequent entries will be excluded. Participant data will be excluded from the analyses if they provide 10 or more identical responses in a row on the CFI. Participants performance on the WCST will not result in any exclusion of data. Descriptive data, missing data and outcome data, as well as any sensitivity analyses that are performed, will be reported in accordance to the STROBE guidelines.[33] Please refer to the document titled "Statistical Analysis Plan (SAP)" (pain-cog-sap.pdf, v03, 2023/02/22) for a full statistical analysis plan.

Sample size and power calculations

Sample size calculations were made under the assumptions of the analysis to be undertaken on the two primary analyses to ensure sufficiently powered primary analyses. Mean and standard deviation estimates of the outcome variables (WCST perseverative responses and CFI scores) in previous studies [34, 8, 16] were utilised. A detectable increase of 5% for the outcome of WCST percentage perseverative responses in the persistent pain group compared to the control group was used, and a detectable decrease of 6 (similarly, 5% of the scale) for the outcome of CFI total score in the persistent pain group compared to the control group was used. A required sample size of 120 was estimated in conjunction with a power of 0.95 and level of

significance of 0.05. The sample size calculations were performed using simulation[35-37] of a beta distributed response of a linear model data generating process as both outcomes are approximately continuous, skewed and bounded. Sensitivity analysis of the required sample size was also performed using alternative model specifications and a range in the number of covariates and their relative effects. The empirical calculations were verified using analytical multiple linear regression power analysis methods available in G*Power 3[38,39] and the 'pwr' package[40] in R (version 4.1.3).[32] Please see the document titled "Sample size calculations for WCST and CFI modelling" (pain-cog-sample-size.pdf, v02, 2022/08/10) for further details. Sample size calculations were undertaken prior to data analysis, but not prior to data collection.

Eligibility Questionnaire

Eligibility Page Instructions:

The first three questions of this survey require you to provide a response to determine whether you are eligible to participate in this study.

Please select the answer that best applies to you:

1. Are you at least 18 years of age?
    a. Yes
    b. No
2. Are you colour-blind?
    a. Yes
    b. No
3. Can you read and understand English?
    a. Yes
    b. No

Summary of Results or Future Research

1. Do you wish to receive a summary of the results for this study?

    a. No

    b. Yes

2. Would you like to be contacted about future research?

    a. No

    b. Yes

If you ticked **yes** to either question above, please leave your email address in the text-box below:

Demographic Information

Demographic Page Instructions:

We would like some information about you and your life. It is not compulsory to answer any of the following questions, but by doing so you allow us to better understand the results of this study.

Please give the answer that best applies to you:

1. What is your age? Please state your age in years _____
2. What sex were you assigned at birth?
    - ☐ Male
    - ☐ Female
    - ☐ Intersex
    - ☐ Not listed here
3. We recognise gender as a social construct and therefore your gender may or may not be the same as your previous answer. What is your current gender identity?

- ☐ Male
- ☐ Female
- ☐ Trans male
- ☐ Trans woman
- ☐ Agender
- ☐ Non-binary
- ☐ My gender identity is not listed. I identify as: _____

4. What is your highest form of education? If currently enrolled, please indicate your highest form of education received so far.
   - ☐ Primary school
   - ☐ Secondary school (i.e., Year 12 or equivalent)
   - ☐ Tertiary education (e.g., Trade qualification/Certificate/Diploma/Bachelor's degree)
   - ☐ University higher degree (e.g., Masters, PhD)
5. How would you best describe your current main activity/occupation?
   - ☐ Unemployed
   - ☐ Working – full time
   - ☐ Working – part time/casual
   - ☐ Unable to work (disability/work cover/carer)
   - ☐ Student
   - ☐ Home duties
   - ☐ Retired
6. Which of the following statements best describes your current financial situation?
   - ☐ Finding it difficult
   - ☐ Getting by
   - ☐ Living comfortably
7. How would you best describe where you live?
   - ☐ Major city
   - ☐ Large town
   - ☐ Small to medium town
   - ☐ Remote community
8. In what country do you live? _____
9. What was your country of birth? _____
10. Do you identify as being a refugee?
    - ☐ Yes
    - ☐ No
11. Do you identify with a particular ethnic or racial group?
    - ☐ No
    - ☐ Yes. If yes, please state what group you identify with: _____

<u>Health information:</u>

Health Information Page Instructions:

We would like some information about your mental and physical health. The following questions are compulsory because they are important for our study.

Please give the answer that best applies to you:

1. Have you previously been diagnosed with a mental health condition?
   - ☐ No
   - ☐ Yes. If yes, please state all that apply: _____
2. Have you previously been diagnosed with any of the following conditions?

- ☐ Stroke
- ☐ Traumatic brain injury
- ☐ Dementia or Alzheimer's disease
- ☐ None of the above

3. In the past week have you taken any of the following medications: Opioids, Antidepressants or Gabapentin (e.g., Lyrica)?
    - ☐ Yes
    - ☐ No

<u>Pain information:</u>

Pain Information Page Instructions:

We would like some information about how you experience pain. The following question(s) are compulsory because they are important for our study.

Please give the answer that best applies to you:

1. For the purposes of this study, we define chronic pain as pain experienced most days for 3 months or longer. According to this definition, do you currently have chronic pain?
    a. Yes
    b. No
2. On a scale from 0 to 10, where **0 = no pain at all and 10 = worse pain imaginable**, how would you rate your average pain in the past week?
3. How long have you been experiencing chronic pain? (Please state the number of months) _____
4. Please select your **worst** area of pain:
    a. Back
    b. Head
    c. Neck
    d. Abdominal
    e. Hip
    f. Pelvic
    g. Hand or arm
    h. Foot or leg
    i. I find it difficult to identify my worst area of pain
5. Where else do you have pain? Please select all that apply:
    a. Back
    b. Head
    c. Neck
    d. Abdominal
    e. Hip
    f. Pelvic
    g. Hand or arm
    h. Foot or leg
    i. I do not have pain anywhere else in my body
6. Have you received a diagnosis for your chronic pain?
    a. No
    b. Yes. If yes, please specify: _____

20-Item Cognitive Flexibility Scale (CFI)

Dennis, J.P & Vander Wal, J.S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research, 34*, 241-253. https://doi.org/10.1007/s10608-009-9276-4

Please use the scale below to indicate the extent to which you agree or disagree with the following statements.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
1	2	3	4	5	6	7

1. I am good at "sizing up" situations.
2. I have a hard time making decisions when faced with difficult situations.
3. I consider multiple options before making a decision.
4. When I encounter difficult situations, I feel like I am losing control.
5. I like to look at difficult situations from many different angles.
6. I seek additional information not immediately available before attributing causes to behavior.
7. When encountering difficult situations, I become so stressed that I can not think of a way to resolve the situation.
8. I try to think about things from another person's point of view.
9. I find it troublesome that there are so many different ways to deal with difficult situations.
10. I am good at putting myself in others' shoes.
11. When I encounter difficult situations, I just don't know what to do.
12. It is important to look at difficult situations from many angles.
13. When in difficult situations, I consider multiple options before deciding how to behave.
14. I often look at a situation from different viewpoints.
15. I am capable of overcoming the difficulties in life that I face.
16. I consider all the available facts and information when attributing causes to behavior.
17. I feel I have no power to change things in difficult situations.
18. When I encounter difficult situations, I stop and try to think of several ways to resolve it.
19. I can think of more than one way to resolve a difficult situation I'm confronted with.
20. I consider multiple options before responding to difficult situations.

Below are the instructions for scoring the measure:

Add up the scores for the items for the Alternatives scale (items 1, 3, 5-6, 8, 10, 12-13, 14,16,18-20) and the Control scale (2, 4, 7, 9, 11, 15 & 17). When adding the items, make sure to reverse score items 2, 4, 7, 9, 11, & 17. Higher scores on both scales are indicative of greater flexibility. You can also use the total score for the CFI too.

There are currently no established cut-off scores for high, moderate, and low flexibility for the CFI total score or the subscale scores. To determine high and low flexibility scores for my own research, I utilized percentile cut-offs as indicated in the attached manuscript.

Hospital Anxiety and Depression Scale (HADS)

Zigmond, A.S & Snaith, R.P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, *67*, 361-370.

Tick the box beside the reply that is closest to how you have been feeling in the past week. Don't take too long over you replies: your immediate is best.

D	A		D	A	
		I feel tense or 'wound up':			**I feel as if I am slowed down:**
	3	Most of the time	3		Nearly all the time
	2	A lot of the time	2		Very often
	1	From time to time, occasionally	1		Sometimes
	0	Not at all	0		Not at all
		I still enjoy the things I used to enjoy:			**I get a sort of frightened feeling like 'butterflies' in the stomach:**
0		Definitely as much		0	Not at all
1		Not quite so much		1	Occasionally
2		Only a little		2	Quite often
3		Hardly at all		3	Very often
		I get a sort of frightened feeling as if something awful is about to happen:			**I have lost interest in my appearance:**
	3	Very definitely and quite badly	3		Definitely
	2	Yes, but not too badly	2		I don't take as much care as I should
	1	A little, but it doesn't worry me	1		I may not take quite as much care
	0	No at all	0		I take just as much care as ever
		I can laugh and see the funny side of things:			**I feel restless as I have to be on the move:**
0		As much as I always could		3	Very much indeed
1		Not quite so much now		2	Quite a lot
2		Definitely not so much now		1	Not very much
3		Not at all		0	Not at all
		Worrying thoughts go through my mind:			**I look forward with enjoyment to things:**
	3	A great deal of the time	0		As much as I ever did
	2	A lot of the time	1		Rather less than I used to
	1	From time to time, but not too often	2		Definitely less than I used to
	0	Only occasionally	3		Hardly at all
		I feel cheerful:			**I get sudden feelings of panic:**
3		Not at all		3	Very often indeed
2		Not often		2	Quite often
1		Sometimes		1	Not very often
0		Most of the time		0	Not at all
		I can sit at ease and feel relaxed:			**I can enjoy a good book or radio or TV program:**
	0	Definitely	0		Often

	1	Usually	1		Sometimes
	2	Not often	2		Not often
	3	Not at all	3		Very seldom

Scoring:
Total score: Depression (D) _____ Anxiety (A)_____
0-7 = Normal
8-10 = Borderline abnormal (borderline case)
11-21 Abnormal (case)

Immune Status Questionnaire (ISQ)

Wilod Versprille, L.J.F., van de Loo, A.J.A.E., Mackus, M., Arnoldy, L., Sulzer, T.A., Vermeulen, S.A., ... & Verster, J.C. (2019). Development and validation of the Immune Status Questionnaire (ISQ). *International Journal of Environmental Research and Public Health*, *16*, 4743. https://doi.org/10.3390/ijerph16234743

Immune Status Questionnaire (ISQ)

Please indicate how often you have had the following complaints **in the past 12 months**:

	Never	Sometimes	Regularly	Often	(Almost) always
Sudden high fever					
Diarrhea					
Headache					
Skin problems (e.g. acne & eczema)					
Muscle and joint pain					
Common Cold					
Coughing					

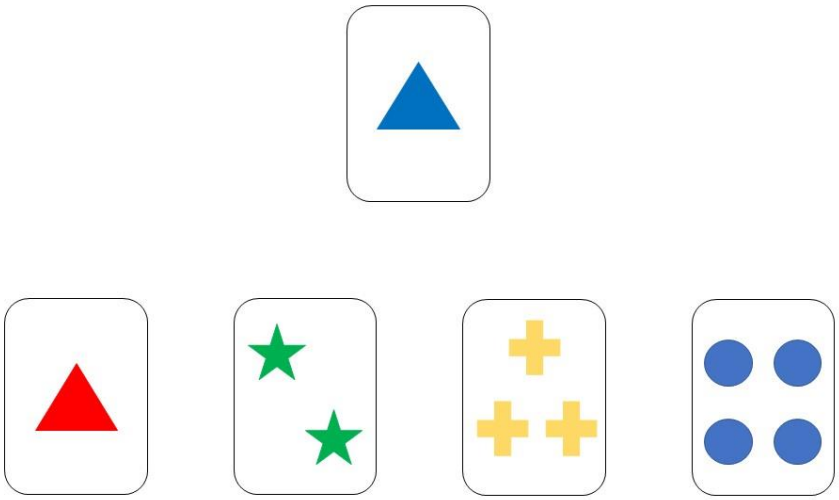
How you feel at this moment?

A	I score my general health the following grade (0 = very bad; 10 = very good)	....	....
B	I score my immune functioning the following grade (0 = very bad; 10 = very good)	....	....
C	Do you have reduced immune functioning at this moment?	Yes	No
D	Do you have a chronic disease? If yes, please specify:	Yes	No

Wisconsin Card Sorting Test (WCST)

Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded.* Psychological Assessment Resources.

Example of the WCST card display:



References

1 Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: Past, present, and future. *Am Psychol* 2014; **69**: 119–30.

2 Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: Scientific advances and future directions. *Psychol Bull* 2007; **133**: 581–624.

3 Goldberg DS, McGee SJ. Pain as a global public health priority. *BMC Public Health* 2011; **11**: 770.

4 Cañas J, Quesada J, Antolí A, Fajardo I. Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics* 2003; **46**: 482–501.

5 Diamond A. Executive functions. *Annu Rev Psychol* 2013; **64**: 135–68.

6 Martin MM, Anderson CM. The cognitive flexibility scale: Three validity studies. *Commun Reports* 1998; **11**: 1–9.

7 Martin MM, Rubin RB. A new measure of cognitive flexibility. *Psychol Rep* 1995; **76**: 623–6.

8 Dennis JP, Vander Wal JS. The cognitive flexibility inventory: Instrument development and estimates of reliability and validity. *Cognit Ther Res* 2010; **34**: 241–53.

9 Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills* 1958; **8**: 271–6.

10 Reitan RM. The relation of the Trail Making Test to organic brain damage. *J Consult Psychol* 1955; **19**: 393–4.

11 Grant DA, Berg E. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J Exp Psychol* 1948; **38**: 404–11.

12 Heaton RK, Chelune GJ, Talley JL, Kay GG, Curtiss G. Wisconsin Card Sorting Test manual: Revised and expanded. Psychological Assessment Resources, 1993.

13 Lounes N, Khan G, Tchanturia K. Assessment of cognitive flexibility in anorexia nervosa - self-report or experimental measure? A brief report. *J Int Neuropsychol Soc* 2011; **17**: 925–8.

14 Miles S, Nedeljkovic M, Phillipou A. Can cognitive flexibility and clinical perfectionism be used to identify people with anorexia nervosa? *J Clin Med* 2022; **11**: 1954.

15 Lee RSC, Hermens DF, Porter MA, Redoblado-Hodge MA. A meta-analysis of cognitive deficits in first-episode major depressive disorder. *J Affect Disord* 2012; **140**: 113–24.

16 Johnco C, Wuthrich VM, Rapee RM. Reliability and validity of two self-report measures of cognitive flexibility. *Psychol Assess* 2014; **26**: 1381–7.

17 Attal N, Masselin-Dubois A, Martinez V, *et al.* Does cognitive functioning predict chronic pain? Results from a prospective surgical cohort. *Brain* 2014; **137**: 904–17.

18 Bell T, Trost Z, Buelow MT, *et al.* Meta-analysis of cognitive performance in fibromyalgia. *J Clin Exp Neuropsychol* 2018; **40**: 698–714.

19 Bunk S, Preis L, Zuidema S, Lautenbacher S, Kunz M. Executive functions and pain. *Zeitschrift für Neuropsychol* 2019; **30**: 169–96.

20 Berryman C, Stanton TR, Bowering KJ, Tabor A, McFarlane A, Moseley GL. Do people with chronic pain have impaired executive function? A meta-analytical review. *Clin Psychol Rev* 2014; **34**: 563–79.

21 Rathbone M, Parkinson W, Rehman Y, Jiang S, Bhandari M, Kumbhare D. Magnitude and variability of effect sizes for the associations between chronic pain and cognitive test performances: A meta-analysis. *Br J Pain* 2016; **10**: 141–55.

22 Gelonch O, Garolera M, Valls J, Rosselló L, Pifarré J. Executive function in fibromyalgia: Comparing subjective and objective measures. *Compr Psychiatry* 2016; **66**: 113–22.

23 Howlett CA, Wewege MA, Berryman C, *et al.* Same room - different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clin Psychol Rev* 2021; **88**: 102061.

24 Howlett CA, Wewege MA, Berryman C, *et al.* Back to the drawing board – the relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts. A systematic review and meta-analysis. *Neuropsychology* 2022; **36**: 347–72.

25 Treede R-D, Rief W, Barke A, *et al.* A classification of chronic pain for ICD-11. *Pain* 2015; **156**: 1003–7.

26 Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983; **67**: 361–70.

27 Wilod Versprille LJF, van de Loo AJAE, Mackus M, *et al.* Development and validation of the Immune Status Questionnaire (ISQ). *Int J Environ Res Public Health* 2019; **16**: 4743.

28 Berg EA. A simple objective technique for measuring flexibility in thinking. *J Gen Psychol* 1948; **39**: 15–22.

29 Tchanturia K, Davies H, Roberts M, *et al.* Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PLoS One* 2012; **7**: e28331.

30 Miles S, Howlett CA, Berryman C, Nedeljkovic M, Moseley GL, Phillipou A. Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behav Res Methods* 2021; **53**: 2083–2091.

31 Millisecond Software. Inquisit 6. 2021. http://www.millisecond.com.

32 R Core Team. R: A language and environment for statistical computing. 2022. http://www.r-project.org/.

33 Vandenbroucke JP, von Elm E, Altman DG, *et al.* Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Int J Surg* 2014; **12**: 1500–24.

34 Greve KW, Stickle TR, Love JM, Bianchini KJ, Stanford MS. Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study. *Arch Clin Neuropsychol* 2005; **20**: 355–64.

35 Arnold BF, Hogan DR, Colford JM, Hubbard AE. Simulation methods to estimate design power: An overview for applied research. *BMC Med Res Methodol* 2011; **11**: 94.

36 Green P, MacLeod C. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol Evol* 2016; **7**: 493–8.

37 Kontopantelis E, Springate DA, Parisi R, Reeves D. Simulation-based power calculations for mixed effects modeling: ipdpower in Stata. *J Stat Softw* 2016; **74**: 1–25.

38 Faul F, Erdfelder E, Lang A-G, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007; **39**: 175–91.

39 Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behav Res Methods* 2009; **41**: 1149–60.

40 Champely S. pwr: Basic functions for power analysis. R package version 1.3-0. 2020. https://cran.r-project.org/package=pwr.

Amendment history

Amendment number	Protocol version number	Date of change	Author(s) of changes	Details of changes
1	Version 2	22/11/2021	CA Howlett	A major strength of Open Sesame is that it is an open-source program, however, the output file of the WCST did not provide information about our primary outcome variables (i.e., perseverative errors and perseverative responses). Thus, we opted to create and administer the online survey via the Inquisit Web software hosted online by Millisecond (Inquisit 6, 2020), which has sufficient data for the WCST outcome variables and in turn, scores the WCST according to the standardised manual (i.e., Heaton et al., 1993). Inquisit also has the added functionality of allowing participants to complete the WCST on a smartphone device. Therefore, we have altered the protocol to reflect this change where appropriate.
2	Version 2	22/11/2021	CA Howlett & GL Moseley	We have made minor changes to the eligibility criteria to improve clarity.
3	Version 2	22/11/2021	CA Howlett, EL Karran, V Bellan & GL Moseley	Major revisions have been made to the demographic questionnaire, including the re-wording of some sections to improve clarity, readability or accuracy – along with the addition and removal of specific questions.
4	Version 2	22/11/2021	CA Howlett	In the *outcomes* section, further details have been provided for the ISQ.
5	Version 2	22/11/2021	CA Howlett	The procedure has been revised to reflect the changes made to the eligibility criteria and the functionality of the Inquisit software.
6	Version 3	10/12/2021	CA Howlett	Changes were made to the response options for question 7 in the “Demographic Information” section of the demographic questionnaire.

7	Version 3	15/12/2021	CA Howlett	Additional response options were added to the 4th and 5th questions of the "Pain Information" section of the demographic questionnaire.
8	Version 3	30/03/2022	CA Howlett	T Stanford was added as a researcher/statistician to this project.
9	Version 3	31/03/2022	CA Howlett & T Stanford	• Revisions were made to the introduction, research questions, eligibility criteria and procedure sections to improve accuracy and clarity. • The subscales of the CFI and WCST perseverative errors were removed as dependent variables for the primary analysis, and only one variable was selected as the primary outcome for each assessment (i.e., CFI total score and WCST % of perseverative responses). • Information that indicates that the HADS has two separate subscales (depression and anxiety) has been added. • *Group* was added as the independent variable; *test order, modality of survey completion, medication use, education level, gender and ISQ scores* are now included as covariates. • The sample size and power calculation has been revised.
10	Version 3	19/07/2022	CA Howlett & T Stanford	The statistical analysis section was revised and now includes reference to a detailed statistical analysis plan.
11	Version 4	13/10/2022	CA Howlett & GL Moseley	Research questions relating to the ISQ were removed because we decided not to proceed with that aspect for the current study.
12	Version 4	13/10/2022	CA Howlett & GL Moseley	We decided not to exclude participants with >10 identical responses in succession on the HADS because we realised that reverse scoring the responses was unnecessary. This is because the Likert scale changes across items – that is, for some items the scale ranges from 0-3 while for other items it ranges from 3-0.

Supplementary 2 (S2): STROBE Statement Checklist

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Title
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Abstract
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Introduction (paragraphs 1-6)
Objectives	3	State specific objectives, including any prespecified hypotheses	Introduction (paragraph 7)
Methods			
Study design	4	Present key elements of study design early in the paper	Materials and methods (paragraph 1)
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Results (paragraph 1)
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	Materials and methods (paragraph 2)
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Materials and methods (paragraphs 3-6); data analysis (paragraphs 4-5)
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	Materials and methods (paragraphs 3-6); data analysis (paragraphs 4-5)
Bias	9	Describe any efforts to address potential sources of bias	Materials and methods (paragraph 8); discussion (paragraph 6); data analysis (paragraph 3)
Study size	10	Explain how the study size was arrived at	Data analysis (paragraph 1)
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable,	Data analysis (paragraphs 2-10)

		describe which groupings were chosen and why	
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Data analysis
		(*b*) Describe any methods used to examine subgroups and interactions	Not applicable
		(*c*) Explain how missing data were addressed	Data analysis (paragraph 3); statistical analysis plan (https://osf.io/9q8ut/).
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	Not applicable
		(*e*) Describe any sensitivity analyses	Data analysis (paragraphs 7-8)
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Figure 2
		(b) Give reasons for non-participation at each stage	Figure 2
		(c) Consider use of a flow diagram	Figure 2
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1
		(b) Indicate number of participants with missing data for each variable of interest	Data analysis (paragraph 3); Table 1
Outcome data	15*	Report numbers of outcome events or summary measures	Tables 1-5
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Data analysis (paragraph 5); Table 2 and 3; data analysis scripts (https://osf.io/9q8ut/);
		(*b*) Report category boundaries when continuous variables were categorized	Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Not applicable

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Results (paragraphs 3-5)
Discussion			
Key results	18	Summarise key results with reference to study objectives	Discussion (paragraphs 1-5)
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Discussion (paragraph 6)
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Discussion (paragraphs 1-6)
Generalisability	21	Discuss the generalisability (external validity) of the study results	Discussion (paragraph 6)
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Declarations

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Supplementary 3 (S3): CHERRIES Checklist

Checklist for Reporting Results of Internet E-Surveys (CHERRIES)

Checklist Item	***Explanation***	***Page Number***
Describe survey design	Describe target population, sample frame. Is the sample a convenience sample? (In "open" surveys this is most likely.)	Materials and methods (paragraph 2)
IRB approval	Mention whether the study has been approved by an IRB.	Materials and methods (paragraph 1)
Informed consent	Describe the informed consent process. Where were the participants told the length of time of the survey, which data were stored and where and for how long, who the investigator was, and the purpose of the study?	Materials and methods (paragraph 7); supplemental file 7
Data protection	If any personal information was collected or stored, describe what mechanisms were used to protect unauthorized access.	Materials and methods (paragraph 8); supplemental file 7
Development and testing	State how the survey was developed, including whether the usability and technical functionality of the electronic questionnaire had been tested before fielding the questionnaire.	Materials and methods (paragraph 9)
Open survey versus closed survey	An "open survey" is a survey open for each visitor of a site, while a closed survey is only open to a sample which the investigator knows (password-protected survey).	Materials and methods (paragraph 7)
Contact mode	Indicate whether or not the initial contact with the potential participants was made on the Internet. (Investigators may also send out questionnaires by mail and allow for Web-based data entry.)	Materials and methods (paragraph 2)
Advertising the survey	How/where was the survey announced or advertised? Some examples are offline media (newspapers), or online (mailing lists – If yes, which ones?) or banner ads (Where were these banner ads posted and what did they look like?). It is important to know the wording of the announcement as it will heavily influence who chooses to participate. Ideally the survey announcement should be published as an appendix.	Materials and methods (paragraph 2); supplemental files 3-6

Web/E-mail	State the type of e-survey (eg, one posted on a Web site, or one sent out through e-mail). If it is an e-mail survey, were the responses entered manually into a database, or was there an automatic method for capturing responses?	Materials and methods (paragraph 2)
Context	Describe the Web site (for mailing list/newsgroup) in which the survey was posted. What is the Web site about, who is visiting it, what are visitors normally looking for? Discuss to what degree the content of the Web site could pre-select the sample or influence the results. For example, a survey about vaccination on a anti-immunization Web site will have different results from a Web survey conducted on a government Web site	Materials and methods (paragraph 2)
Mandatory/voluntary	Was it a mandatory survey to be filled in by every visitor who wanted to enter the Web site, or was it a voluntary survey?	Materials and methods (paragraph 7)
Incentives	Were any incentives offered (eg, monetary, prizes, or non-monetary incentives such as an offer to provide the survey results)?	Materials and methods (paragraph 9)
Time/Date	In what timeframe were the data collected?	Results (paragraph 1)
Randomization of items or questionnaires	To prevent biases items can be randomized or alternated.	Materials and methods (paragraph 8)
Adaptive questioning	Use adaptive questioning (certain items, or only conditionally displayed based on responses to other items) to reduce number and complexity of the questions.	Materials and methods (paragraph 5)
Number of Items	What was the number of questionnaire items per page? The number of items is an important factor for the completion rate.	Materials and methods (paragraph 8)
Number of screens (pages)	Over how many pages was the questionnaire distributed? The number of items is an important factor for the completion rate.	Materials and methods (paragraph 8)
Completeness check	It is technically possible to do consistency or completeness checks before the questionnaire is submitted. Was this done, and if “yes”, how (usually JAVAScript)? An alternative is to check for completeness after the questionnaire has been submitted (and highlight mandatory items). If this has been done, it should be reported. All items should provide a non-response option such as “not applicable” or “rather not say”, and selection of one response option should be enforced.	Materials and methods (paragraph 8); supplemental file 9
Review step	State whether respondents were able to review and change their answers (eg, through a Back button or a Review step which	Materials and methods

	displays a summary of the responses and asks the respondents if they are correct).	(paragraph 8)
Unique site visitor	If you provide view rates or participation rates, you need to define how you determined a unique visitor. There are different techniques available, based on IP addresses or cookies or both.	Not applicable
View rate (Ratio of unique survey visitors/unique site visitors)	Requires counting unique visitors to the first page of the survey, divided by the number of unique site visitors (not page views!). It is not unusual to have view rates of less than 0.1 % if the survey is voluntary.	Not applicable
Participation rate (Ratio of unique visitors who agreed to participate/unique first survey page visitors)	Count the unique number of people who filled in the first survey page (or agreed to participate, for example by checking a checkbox), divided by visitors who visit the first page of the survey (or the informed consents page, if present). This can also be called "recruitment" rate.	Not applicable
Completion rate (Ratio of users who finished the survey/users who agreed to participate)	The number of people submitting the last questionnaire page, divided by the number of people who agreed to participate (or submitted the first survey page). This is only relevant if there is a separate "informed consent" page or if the survey goes over several pages. This is a measure for attrition. Note that "completion" can involve leaving questionnaire items blank. This is not a measure for how completely questionnaires were filled in. (If you need a measure for this, use the word "completeness rate".)	Not applicable
Cookies used	Indicate whether cookies were used to assign a unique user identifier to each client computer. If so, mention the page on which the cookie was set and read, and how long the cookie was valid. Were duplicate entries avoided by preventing users access to the survey twice; or were duplicate database entries having the same user ID eliminated before analysis? In the latter case, which entries were kept for analysis (eg, the first entry or the most recent)?	Not applicable
IP check	Indicate whether the IP address of the client computer was used to identify potential duplicate entries from the same user. If so, mention the period of time for which no two entries from the same IP address were allowed (eg, 24 hours). Were duplicate entries avoided by preventing users with the same IP address access to the survey twice; or were duplicate database entries having the same IP address within a given period of time eliminated before analysis? If the latter, which entries were kept for analysis (eg, the first entry or the most recent)?	Not applicable
Log file analysis	Indicate whether other techniques to analyze the log file for identification of multiple entries were used. If so, please describe.	Data analysis (paragraph 3)

Registration	In “closed” (non-open) surveys, users need to login first and it is easier to prevent duplicate entries from the same user. Describe how this was done. For example, was the survey never displayed a second time once the user had filled it in, or was the username stored together with the survey results and later eliminated? If the latter, which entries were kept for analysis (eg, the first entry or the most recent)?	Not applicable
Handling of incomplete questionnaires	Were only completed questionnaires analyzed? Were questionnaires which terminated early (where, for example, users did not go through all questionnaire pages) also analyzed?	Data analysis (paragraph 3) supplemental file 7
Questionnaires submitted with an atypical timestamp	Some investigators may measure the time people needed to fill in a questionnaire and exclude questionnaires that were submitted too soon. Specify the timeframe that was used as a cut-off point, and describe how this point was determined.	Not applicable
Statistical correction	Indicate whether any methods such as weighting of items or propensity scores have been used to adjust for the non-representative sample; if so, please describe the methods.	Not applicable

This checklist has been modified from Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res. 2004 Sep 29;6(3):e34 [erratum in J Med Internet Res. 2012; 14(1): e8.]. Article available at https://www.jmir.org/2004/3/e34/; erratum available https://www.jmir.org/2012/1/e8/. Copyright ©Gunther Eysenbach. Originally published in the Journal of Medical Internet Research, 29.9.2004 and 04.01.2012.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/2.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited.

Supplementary 4 (S4): Recruitment flyer



Supplementary 5 (S5): Advertisements for social media platforms and websites

Researchers from the University of South Australia are looking for individuals with or without persistent pain to participate in an online study investigating cognitive performance and pain.

If you are pain-free or living with persistent pain (i.e., pain experienced most days for 3 months or longer), at least 18 years of age, able to read and understand English, have normal or corrected to normal (e.g., glasses or contact lenses) vision, have access to a computer/laptop or smartphone device with internet access and are not colour-blind then you are considered to be eligible for this study.

Participation involves the completion of an online survey that takes approximately 20-30 minutes. If you are not eligible to participate but you know someone who might be, please feel free to share this post with them.

To find out more about this study or to participate in this research, please visit the following link: [insert survey link]. Alternatively, you can contact the primary researcher of this study: Caitlin Howlett (caitlin.howlett@mymail.unisa.edu.au).




University of
South Australia
Can you help us to better understand persistent pain?
Follow the link in the post to participate in the online survey
Adults who are at least 18 years of age WITH or WITHOUT persistent pain can take part in this research
*Please note that other eligibility criteria apply - please refer to the details outlined in the post for more information
This project has been approved by the University of South Australia's Human Research Ethics Committee (Protocol No. 202439)

Supplementary 6 (S6): Post for UniSA's Research Volunteers website

Title of study:	Understanding the role of cognitive performance in persistent pain: an online study
About the study:	This study aims to understand the role of cognitive performance in persistent pain. It is estimated that 1 in 5 people experience pain that persists for longer than 3 months, making it a global health concern. Understanding what makes some people more susceptible to the development and maintenance of persistent pain is important for determining therapeutic targets that may aid in recovery. The survey takes approximately 20-30 minutes to complete. To complete this online survey, you will be required to download a small file onto your computer or smartphone; the file will allow you to install the Inquisit software (hosted online by Millisecond), which is the program used to run our survey. This file can be deleted after you have completed the study and instructions on how to remove the file from your device are provided at the end of the survey.
Eligibility requirements:	• At least 18 years of age • Able to read and understand English • Pain-free or living with persistent pain (i.e., pain experienced most days for 3 months or longer) • Have access to a computer/laptop or smartphone device with internet access • Have normal or corrected to normal (i.e., glasses or contact lenses) vision • Are not colour-blind • Location: Worldwide
Benefits:	There will be no direct benefits to you, but this study will provide important knowledge that will assist us in understanding potential targets for treatment for people experiencing persistent pain.
Link to the survey:	Get involved in this study by following this link: https://mili2nd.co/ba5b This project has been approved by the University of South Australia's Human Research Ethics Committee (Protocol 202439). If you have any questions or concerns before or after the experiment, or would like further information, please do not hesitate to contact the primary investigator of this project: Caitlin Howlett Email: caitlin.howlett@mymail.unisa.edu.au

Supplementary 7 (S7): Recruitment email for the Research Participation Sign-up System



PARTICIPANT EMAIL RECRUITMENT SCRIPT

Dear XXXX

You are invited to participate in a study investigating the role of cognitive performance in persistent pain.

This study is seeking volunteers who:

- Are at least 18 years of age
- Able to read and understand English
- Pain-free or living with persistent pain (i.e., pain experienced most days for 3 months or longer)
- Have access to a computer/laptop or smartphone device with internet access
- Have normal or corrected to normal (e.g., glasses or contact lenses) vision
- Are not colour-blind

If you are interested in participating you can follow the link to the online survey here: https://mili2nd.co/ba5b

If you would like to know more about the study, we have provided a detailed description below of the purpose of the research, what participation in this study will involve and the management of the data.

If you have any questions or concerns before or after the experiment, please do not hesitate to contact PhD candidate Caitlin Howlett via email at caitlin.howlett@mymail.unisa.edu.au or via phone on 8302 2454.

If you are not eligible but know someone who might be, please feel free to forward this email to them.

We look forward to hearing from you.

Yours sincerely,

Caitlin Howlett, BPsych(Hons)
PhD candidate, Body in Mind Research Group, IIMPACT in Health, Allied Health & Human Performance,

UniSA
Pronouns: she/her/hers
Email: caitlin.howlett@mymail.unisa.edu.au

This project has been approved by the University of South Australia's Human Research Ethics Committee (Protocol No. 202439). If you have any ethical concerns about the project or questions about your rights as a participant, please contact the Executive Officer of this Committee (Email: humanethics@unisa.edu.au; Tel: 8302 6330)

What is the study about?

It is estimated that 1 in 5 people experience pain that persists for longer than 3 months, making it a global health concern. Understanding what makes some people more susceptible to the development and maintenance of persistent pain is important for determining therapeutic targets that may aid in recovery. This study is an online survey that aims to understand the role of cognitive performance in persistent pain.

Who can participate?

You are eligible to participate if you are at least 18 years of age, able to read and understand English, are pain-free or living with persistent pain (i.e., pain experienced most days for 3 months or longer), have access to a computer/laptop or smartphone device with internet access, have normal or corrected to normal (e.g., glasses or contact lenses) vision, and are not colour-blind. You are not required to be experiencing persistent pain to participate.

Do I have to participate?

No. As with all research at the University of South Australia, your involvement in this study is voluntary. You are also free to withdraw from the study at any time and it won't affect your relationship with the researchers or your relationship with the University of South Australia now or in the future. Any information collected prior to your withdrawal will not be analysed. If you complete the survey but later decide that you wish to withdraw from the study, you will have one week after completing the survey to contact the researcher to withdraw from the study and request that information collected from you is not included in our analysis. Please note that any information collected until the point that you withdraw will be retained, unless you explicitly request to have this excluded.

What will participation in this study involve?

If you agree to participate, you will be required to complete an online screening questionnaire that will determine whether you are eligible to participate. If you are deemed eligible, you will then be asked to fill in an online questionnaire that includes questions surrounding your age, gender, education level, basic health background and experience of pain. You will then be provided with three additional short questionnaires, two asking you to how much you agree or disagree with various statements and the other asking you to reflect on how often you have experienced health complaints (e.g., headaches) in the past 12 months. Lastly, you will be asked to complete a self-paced card-sorting task. This online survey is entirely anonymous and will take approximately 20-30 minutes to complete; however, please be mindful that completion times can vary since there is no time limit for the card-sorting task. To complete this online survey, you will be required to download a small file onto your computer or smartphone; the file will allow you to install the Inquisit software (hosted online by Millisecond), which is the program used to run our survey. This

file can be deleted after you have completed the study and instructions on how to remove the file from your device are provided at the end of the survey. Millisecond is a secure program that stores information anonymously so that it cannot be traced back to any individual's identity. Because the data are anonymous, Millisecond is unable to provide individuals access to their data. Any data collected by Millisecond is stored at a fixed data centre and Millisecond will never transfer data to a third-party without written permission from the customer (i.e., the researchers). Millisecond automatically logs information from your internet browser such as IP addresses and the type of internet browser for security purposes (i.e., detecting and preventing fraud, unauthorised access, denial of service, and other malicious attacks); however, this information is deleted after 3 months. Millisecond complies with the EU-U.S. Privacy Shield Framework as set forth by the U.S. Department of Commerce.

What are the risks and benefits in participating?

There will be no direct benefits to you, but this study will provide important knowledge that will assist us in understanding potential targets for treatment for people experiencing persistent pain. The risks associated with this study are minimal but some people may find the content within this survey distressing, particularly when responding to questions relating to mental/physical health. Due to the nature of this study, some information will be withheld from you until the end of your involvement. However, there is no added risk of harm or any reason to suspect you would not consent if you had all the details at the start. After testing is completed, you will be given a full explanation of the study, including reasons for concealing some details. If at any point you feel distressed, you are welcome to stop the survey without penalty. Doing so will not affect your relationship with the researchers of the University of South Australia. If you require further support, please contact your local general practitioner, psychologist/counsellor or support hotline (e.g., Lifeline or Beyond Blue).

Who will be given access to my information and how will it remain confidential?

All the information you provide will be de-identified prior to data analysis. Electronic data will be stored on a secure sever dedicated for the survey software. Survey data will be saved in a password protected excel spreadsheet and stored on a secure UniSA server indefinitely or for a minimum of five years as required by South Australian legislation. The researcher will take every care to remove any identifying material from the responses you provide as early as possible. Likewise, individuals' responses will be kept confidential by the researcher unless required by law, and participants will not be identified by the reporting of the research. However, the researcher cannot guarantee the confidentiality or anonymity of material transferred by email or the internet. Any non-identifiable data that is collected has the potential to be used for future research/secondary analyses and be uploaded to Open Science Framework – an open access data sharing platform – to improve transparency and reproducibility of scientific research, for which ethical approval will be sought.

Will I receive a summary of the results of the research?

The results will be published in a doctorate thesis and in a scientific journal, as well as having the potential to be presented at a conference. If you wish to receive a summary of the results, you will have the option to provide your preferred email address. Once the findings have been formally published, we will send this information to you via email.

Supplementary 8 (S8): Information sheet and consent form



PARTICIPANT INFORMATION SHEET & CONSENT FORM

Understanding the role of cognitive performance in persistent pain

Caitlin Howlett, Dr. Tyman Stanford, Dr. Scott Coussens, Dr. Carolyn Berryman, Dr. Emma L. Karran, Dr. Valeria Bellan, Dr. Stephanie Miles, Prof. Lorimer Moseley

Thank you for considering participating in this research study, which is being conducted by the Body in Mind Research Group within the Innovation, Implementation and Clinical Translation (IIMPACT) in Health Group at the University of South Australia. This information sheet is intended to provide you with sufficient information to make an informed decision about participating in this study. Please read this page carefully in order to determine whether you are indeed eligible to take part.

What is the study about?
It is estimated that 1 in 5 people experience pain that persists for longer than 3 months, making it a global health concern. Understanding what makes some people more susceptible to the development and maintenance of persistent pain is important for determining therapeutic targets that may aid in recovery. This study is an online survey that aims to understand the role of cognitive performance in persistent pain.

Who can participate?
You are eligible to participate if you are at least 18 years of age, able to read and understand English, are pain-free or living with persistent pain (i.e., pain experienced most days for 3 months or longer), have access to a computer/laptop or smartphone device with internet access, have normal or corrected to normal (e.g., glasses or contact lenses) vision, and are not colour-blind. You are not required to be experiencing persistent pain to participate.

Do I have to participate?
No. As with all research at the University of South Australia, your involvement in this study is voluntary. You are also free to withdraw from the study at any time and it won't affect your relationship with the researchers or your relationship with the University of South Australia now or in the future. Any information collected prior to your withdrawal will not be analysed. If you complete the survey but later decide that you wish to withdraw from the study, you will have one week after completing the survey to contact the researcher to withdraw from the study and request that information collected from you is not included in our analysis. Please note that any information collected until the point that you withdraw will be retained, unless you explicitly request to have this excluded.

What will participation in this study involve?

If you agree to participate, you will be required to complete an online screening questionnaire that will determine whether you are eligible to participate. If you are deemed eligible, you will then be asked to fill in an online questionnaire that includes questions surrounding your age, gender, education level, basic health background and experience of pain. You will then be provided with three additional short questionnaires, two asking you to how much you agree or disagree with various statements and the other asking you to reflect on how often you have experienced health complaints (e.g., headaches) in the past 12 months. Lastly, you will be asked to complete a self-paced card-sorting task. This online survey is entirely anonymous and will take approximately 20-30 minutes to complete; however, please be mindful that completion times can vary since there is no time limit for the card-sorting task. To complete this online survey, you will be required to download a small file onto your computer or smartphone; the file will allow you to install the Inquisit software (hosted online by Millisecond), which is the program used to run our survey. This file can be deleted after you have completed the study and instructions on how to remove the file from your device are provided at the end of the survey. Millisecond is a secure program that stores information anonymously so that it cannot be traced back to any individual's identity. Because the data are anonymous, Millisecond is unable to provide individuals access to their data. Any data collected by Millisecond is stored at a fixed data centre and Millisecond will never transfer data to a third-party without written permission from the customer (i.e., the researchers). Millisecond automatically logs information from your internet browser such as IP addresses and the type of internet browser for security purposes (i.e., detecting and preventing fraud, unauthorised access, denial of service, and other malicious attacks); however, this information is deleted after 3 months. Millisecond complies with the EU-U.S. Privacy Shield Framework as set forth by the U.S. Department of Commerce.

What are the risks and benefits in participating?

There will be no direct benefits to you, but this study will provide important knowledge that will assist us in understanding potential targets for treatment for people experiencing persistent pain. The risks associated with this study are minimal but some people may find the content within this survey distressing, particularly when responding to questions relating to mental/physical health. Due to the nature of this study, some information will be withheld from you until the end of your involvement. However, there is no added risk of harm or any reason to suspect you would not consent if you had all the details at the start. After testing is completed, you will be given a full explanation of the study, including reasons for concealing some details. If at any point you feel distressed, you are welcome to stop the survey without penalty. Doing so will not affect your relationship with the researchers of the University of South Australia. If you require further support, please contact your local general practitioner, psychologist/counsellor or support hotline (see below).

Emergency hotline contact information	
Australia	Beyond Blue: 1300 22 4636 Lifeline: 13 11 14
Canada	Crisis Services Canda: 1-833-456-4566 (toll free)
United States of America	Mental Health America: 1-800-273-TALK (8255)
United Kingdom	HopeLine UK: 0800 068 4141
Republic of Ireland	Lifeline: 0808 808 8000

New Zealand	Lifeline New Zealand: 09 5222 999 (Auckland) or 0800 542 354 (Rest of NZ)
Europe	France SOS Help (English speaking): 01 46 21 46 46 Germany Telefonseelsorge Deutschland (German speaking): 0800-111 0 222 International Helpline Berlin (English speaking): 030-44 01 06 07 Italy Samaritans: 800 86 00 22
South Africa	Lifeline Southern Africa: 0861 322 322
Japan	Befrienders International, Tokyo: +81 (0) 3 5286 9090 Tokyo English Lifeline: 107-0062

Who will be given access to my information and how will it remain confidential?
All the information you provide will be de-identified prior to data analysis. Electronic data will be stored on a secure sever dedicated for the survey software. Survey data will be saved in a password protected excel spreadsheet and stored on a secure UniSA server indefinitely or for a minimum of five years as required by South Australian legislation. The researcher will take every care to remove any identifying material from the responses you provide as early as possible. Likewise, individuals' responses will be kept confidential by the researcher unless required by law, and participants will not be identified by the reporting of the research. However, the researcher cannot guarantee the confidentiality or anonymity of material transferred by email or the internet. Any non-identifiable data that is collected has the potential to be used for future research/secondary analyses and be uploaded to Open Science Framework – an open access data sharing platform – to improve transparency and reproducibility of scientific research, for which ethical approval will be sought.

Will I receive a summary of the results of the research?
The results will be published in a doctorate thesis and in a scientific journal, as well as having the potential to be presented at a conference. If you wish to receive a summary of the results, you will have the option to provide your preferred email address. Once the findings have been formally published, we will send this information to you via email.

Further information and contact details
If you have any questions or concerns before or after the experiment, or would like further information, please do not hesitate to contact the primary investigator of this project:

Caitlin Howlett

Email: caitlin.howlett@mymail.unisa.edu.au

Tel: +618 8302 2454

This project has been approved by the University of South Australia's Human Research Ethics Committee (Protocol No. 202439). If you have any ethical concerns about the project or questions about your rights as a participant, or should you or any third parties wish to lodge a complaint

about either the study or the way it is being conducted, please feel free to contact the Executive Officer of UniSA's Human Research Ethics Committee:

Executive Officer, UniSA Human Research Ethics Committee

Email: humanethics@unisa.edu.au

Tel: +618 8302 6330

By completing and submitting the survey, you are indicating that:

- You have read and understood the Participant Information Sheet and give your consent to be involved in the research
- You understand the nature of your involvement in the project
- You understand that there is a slight risk of psychological/emotional distress whilst disclosing information relating to your mental/physical health
- You understand that you may not directly benefit from taking part in the project
- You understand that you can withdraw your consent at any time (during or within one week of completing the study) and that this will not affect your status now or in the future
- You understand that while information gained during the project may be published, you will not be identified and your personal results will remain confidential unless required by law
- You give the researcher permission to use your non-identifiable data for the purposes of future research/secondary analyses and sharing data (ethical approval will be sought)
- You understand that you may be contacted about future research if you provide an email address
- You meet the eligibility requirements:
    - At least 18 years of age
    - Able to read and understand English
    - Have access to a computer/laptop or smartphone device with internet access
    - Have normal or corrected to normal (e.g., glasses or contact lenses) vision
    - Are not colour-blind

Please note that your responses will be stored separately to your personal information (i.e., email address) so your information is kept confidential.

By continuing the survey, you are indicating that you have read the Participant Information Sheet, understand the nature and purpose of the research project and agree to participate.

Supplementary 9 (S9): Debriefing form/completion page

University of South Australia

PARTICIPANT DEBRIEFING FORM/COMPLETION PAGE

Thank you for participating in this study. Your participation is greatly appreciated. This form contains further information about the purpose of the research project, instructions for how to uninstall Inquisit from your device, details for mental health services should you require support, and contact information if you would like to get in touch with the primary investigator of this project.

Purpose of the research study:

Persistent pain is a common health problem affecting 1 in 5 adults. We previously informed you that the purpose of the study was to understand the role of cognitive performance in persistent pain; however, the aim of our research is to understand the role of ***cognitive flexibility*** in persistent pain. Specifically, we are interested in determining whether there are differences between people with and without persistent pain in neurocognitive and self-reported cognitive flexibility. The information collected in this study will allow us to better understand the factors that make some people more likely to develop persistent pain, and why it might be harder for some people to recover. The more we know about persistent pain and the factors that contribute to such conditions, the better the treatment and preventative strategies. In order to adequately test our research question, we could not disclose certain information about the purpose of the study and we hope you understand why this was necessary. If you decide that you wish to withdraw from the study, you will have one week after completing the survey to contact the researcher to withdraw from the study and request that information collected from you is not included in our analysis (refer to the contact details below). Please do not disclose the goal of the research to anyone who might participate in this study in the future as this could impact the results of the study.

Uninstalling Inquisit Web on Windows/Mac/iOS/Android:

For instructions on how to uninstall Inquisit Web from your device, follow this link: https://mili2nd.co/web/uninstall

Mental health support:

We acknowledge that some people may have found the content within this survey distressing, particularly when responding to questions relating to mental/physical health. We want to ensure that you are aware of resources that are available to you, should you decide you need assistance

at any time. If you require further support, please contact your local general practitioner, psychologist/counsellor or support hotline (see below).

Emergency hotline contact information	
Australia	Beyond Blue: 1300 22 4636 Lifeline: 13 11 14
Canada	Crisis Services Canda: 1-833-456-4566 (toll free)
United States of America	Mental Health America: 1-800-273-TALK (8255)
United Kingdom	HopeLine UK: 0800 068 4141
Republic of Ireland	Lifeline: 0808 808 8000
New Zealand	Lifeline New Zealand: 09 5222 999 (Auckland) or 0800 542 354 (Rest of NZ)
Europe	France SOS Help (English speaking): 01 46 21 46 46 Germany Telefonseelsorge Deutschland (German speaking): 0800-111 0 222 International Helpline Berlin (English speaking): 030-44 01 06 07 Italy Samaritans: 800 86 00 22
South Africa	Lifeline Southern Africa: 0861 322 322
Japan	Befrienders International, Tokyo: +81 (0) 3 5286 9090 Tokyo English Lifeline: 107-0062

Contact information:

If you have any questions or concerns about this study, please do not hesitate to contact the primary investigator of this project, Caitlin Howlett via email (caitlin.howlett@mymail.unisa.edu.au) or phone (+618 8302 2454).

This project has been approved by the University of South Australia's Human Research Ethics Committee (Protocol No. 202439). If you have any ethical concerns about the project or questions about your rights as a participant, or should you or any third parties wish to lodge a complaint about either the study or the way it is being conducted, please feel free to contact the Executive Officer of UniSA's Human Research Ethics Committee (HREC) via email (humanethics@unisa.edu.au) or phone (+618 8302 6330).

Supplementary 10 (S10): Online survey

Eligibility Questionnaire

Eligibility Page Instructions:

The first three questions of this survey require you to provide a response to determine whether you are eligible to participate in this study.

Please select the answer that best applies to you:

4. Are you at least 18 years of age?
    a. Yes
    b. No
5. Are you colour-blind?
    a. Yes
    b. No
6. Can you read and understand English?
    a. Yes
    b. No

Summary of Results or Future Research

3. Do you wish to receive a summary of the results for this study?

    c. No

    d. Yes

4. Would you like to be contacted about future research?

    a. No

    b. Yes

If you ticked **yes** to either question above, please leave your email address in the text-box below:

Demographic Information

Demographic Page Instructions:

We would like some information about you and your life. It is not compulsory to answer any of the following questions, but by doing so you allow us to better understand the results of this study.

Please give the answer that best applies to you:

12. What is your age? Please state your age in years _____
13. What sex were you assigned at birth?
    - Male
    - Female
    - Intersex
    - Not listed here
14. We recognise gender as a social construct and therefore your gender may or may not be the same as your previous answer. What is your current gender identity?

- Male
- Female
- Trans male
- Trans woman
- Agender
- Non-binary
- My gender identity is not listed. I identify as: _____

15. What is your highest form of education? If currently enrolled, please indicate your highest form of education received so far.
    - Primary school
    - Secondary school (i.e., Year 12 or equivalent)
    - Tertiary education (e.g., Trade qualification/Certificate/Diploma/Bachelor's degree)
    - University higher degree (e.g., Masters, PhD)
16. How would you best describe your current main activity/occupation?
    - Unemployed
    - Working – full time
    - Working – part time/casual
    - Unable to work (disability/work cover/carer)
    - Student
    - Home duties
    - Retired
17. Which of the following statements best describes your current financial situation?
    - Finding it difficult
    - Getting by
    - Living comfortably
18. How would you best describe where you live?
    - Major city
    - Large town
    - Small to medium town
    - Remote community
19. In what country do you live? _____
20. What was your country of birth? _____
21. Do you identify as being a refugee?
    - Yes
    - No
22. Do you identify with a particular ethnic or racial group?
    - No
    - Yes. If yes, please state what group you identify with: _____

<u>Health information:</u>

Health Information Page Instructions:

We would like some information about your mental and physical health. The following questions are compulsory because they are important for our study.

Please give the answer that best applies to you:

4. Have you previously been diagnosed with a mental health condition?
    - No
    - Yes. If yes, please state all that apply: _____
5. Have you previously been diagnosed with any of the following conditions?
    - Stroke
    - Traumatic brain injury
    - Dementia or Alzheimer's disease

- None of the above

6. In the past week have you taken any of the following medications: Opioids, Antidepressants or Gabapentin (e.g., Lyrica)?
   - Yes
   - No

Pain information:

Pain Information Page Instructions:

We would like some information about how you experience pain. The following question(s) are compulsory because they are important for our study.

Please give the answer that best applies to you:

7. For the purposes of this study, we define chronic pain as pain experienced most days for 3 months or longer. According to this definition, do you currently have chronic pain?
   a. Yes
   b. No
8. On a scale from 0 to 10, where **0 = no pain at all and 10 = worse pain imaginable**, how would you rate your average pain in the past week?
9. How long have you been experiencing chronic pain? (Please state the number of months) _____
10. Please select your **worst** area of pain:
    a. Back
    b. Head
    c. Neck
    d. Abdominal
    e. Hip
    f. Pelvic
    g. Hand or arm
    h. Foot or leg
    i. I find it difficult to identify my worst area of pain
11. Where else do you have pain? Please select all that apply:
    a. Back
    b. Head
    c. Neck
    d. Abdominal
    e. Hip
    f. Pelvic
    g. Hand or arm
    h. Foot or leg
    i. I do not have pain anywhere else in my body
12. Have you received a diagnosis for your chronic pain?
    a. No
    b. Yes. If yes, please specify: _____

20-Item Cognitive Flexibility Scale (CFI)

Dennis, J.P & Vander Wal, J.S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research, 34*, 241-253. https://doi.org/10.1007/s10608-009-9276-4

Please use the scale below to indicate the extent to which you agree or disagree with the following statements.

Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
1	2	3	4	5	6	7

21. I am good at “sizing up” situations.
22. I have a hard time making decisions when faced with difficult situations.
23. I consider multiple options before making a decision.
24. When I encounter difficult situations, I feel like I am losing control.
25. I like to look at difficult situations from many different angles.
26. I seek additional information not immediately available before attributing causes to behavior.
27. When encountering difficult situations, I become so stressed that I can not think of a way to resolve the situation.
28. I try to think about things from another person’s point of view.
29. I find it troublesome that there are so many different ways to deal with difficult situations.
30. I am good at putting myself in others’ shoes.
31. When I encounter difficult situations, I just don’t know what to do.
32. It is important to look at difficult situations from many angles.
33. When in difficult situations, I consider multiple options before deciding how to behave.
34. I often look at a situation from different viewpoints.
35. I am capable of overcoming the difficulties in life that I face.
36. I consider all the available facts and information when attributing causes to behavior.
37. I feel I have no power to change things in difficult situations.
38. When I encounter difficult situations, I stop and try to think of several ways to resolve it.
39. I can think of more than one way to resolve a difficult situation I’m confronted with.
40. I consider multiple options before responding to difficult situations.

Below are the instructions for scoring the measure:

Add up the scores for the items for the Alternatives scale (items 1, 3, 5-6, 8, 10, 12-13, 14,16,18-20) and the Control scale (2, 4, 7, 9, 11, 15 & 17). When adding the items, make sure to reverse score items 2, 4, 7, 9, 11, & 17. Higher scores on both scales are indicative of greater flexibility. You can also use the total score for the CFI too.

There are currently no established cut-off scores for high, moderate, and low flexibility for the CFI total score or the subscale scores. To determine high and low flexibility scores for my own research, I utilized percentile cut-offs as indicated in the attached manuscript.

Hospital Anxiety and Depression Scale (HADS)

Zigmond, A.S & Snaith, R.P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, *67*, 361-370.

Tick the box beside the reply that is closest to how you have been feeling in the past week. Don't take too long over you replies: your immediate is best.

D	A		D	A	
		I feel tense or 'wound up':			**I feel as if I am slowed down:**
	3	Most of the time	3		Nearly all the time
	2	A lot of the time	2		Very often
	1	From time to time, occasionally	1		Sometimes
	0	Not at all	0		Not at all
		I still enjoy the things I used to enjoy:			**I get a sort of frightened feeling like 'butterflies' in the stomach:**
0		Definitely as much		0	Not at all
1		Not quite so much		1	Occasionally
2		Only a little		2	Quite often
3		Hardly at all		3	Very often
		I get a sort of frightened feeling as if something awful is about to happen:			**I have lost interest in my appearance:**
	3	Very definitely and quite badly	3		Definitely
	2	Yes, but not too badly	2		I don't take as much care as I should
	1	A little, but it doesn't worry me	1		I may not take quite as much care
	0	No at all	0		I take just as much care as ever
		I can laugh and see the funny side of things:			**I feel restless as I have to be on the move:**
0		As much as I always could		3	Very much indeed
1		Not quite so much now		2	Quite a lot
2		Definitely not so much now		1	Not very much
3		Not at all		0	Not at all
		Worrying thoughts go through my mind:			**I look forward with enjoyment to things:**
	3	A great deal of the time	0		As much as I ever did
	2	A lot of the time	1		Rather less than I used to
	1	From time to time, but not too often	2		Definitely less than I used to
	0	Only occasionally	3		Hardly at all
		I feel cheerful:			**I get sudden feelings of panic:**
3		Not at all		3	Very often indeed
2		Not often		2	Quite often
1		Sometimes		1	Not very often
0		Most of the time		0	Not at all
		I can sit at ease and feel relaxed:			**I can enjoy a good book or radio or TV program:**

	0	Definitely	0		Often
	1	Usually	1		Sometimes
	2	Not often	2		Not often
	3	Not at all	3		Very seldom

Scoring:
Total score: Depression (D) _____ Anxiety (A)_____
0-7 = Normal
8-10 = Borderline abnormal (borderline case)
11-21 Abnormal (case)

Immune Status Questionnaire (ISQ)

Wilod Versprille, L.J.F., van de Loo, A.J.A.E., Mackus, M., Arnoldy, L., Sulzer, T.A.L., Vermeulen, S.A., ... & Verster, J.C. (2019). Development and validation of the Immune Status Questionnaire (ISQ). *International Journal of Environmental Research and Public Health*, *16*, 4743. https://doi.org/10.3390/ijerph16234743

Immune Status Questionnaire (ISQ)

Please indicate how often you have had the following complaints **in the past 12 months**:

	Never	Sometimes	Regularly	Often	(Almost) always
Sudden high fever					
Diarrhea					
Headache					
Skin problems (e.g. acne & eczema)					
Muscle and joint pain					
Common Cold					
Coughing					

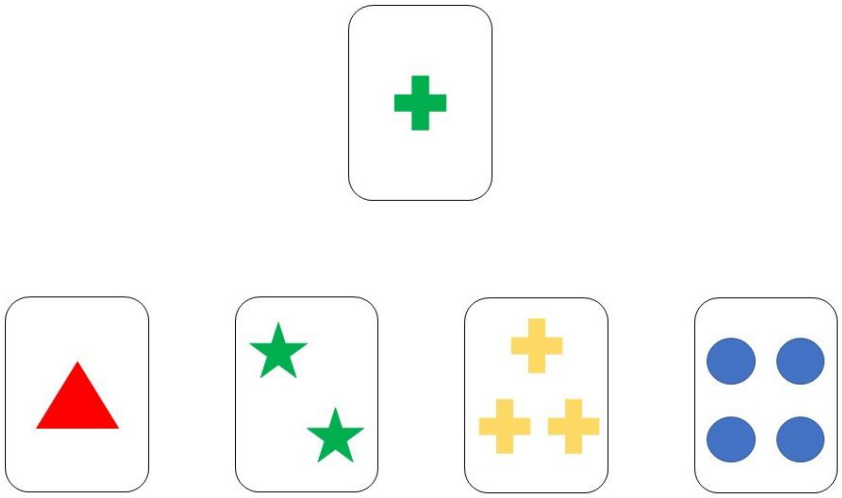
How you feel at this moment?

A	I score my general health the following grade (0 = very bad; 10 = very good)	....	....
B	I score my immune functioning the following grade (0 = very bad; 10 = very good)	....	....
C	Do you have reduced immune functioning at this moment?	Yes	No
D	Do you have a chronic disease? If yes, please specify:	Yes	No

Wisconsin Card Sorting Test (WCST)

Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded.* Psychological Assessment Resources.

Example of the WCST card display:



Supplementary 11 (S11): Participant demographics (continued)

Participant Characteristics (N = 196).

Characteristic	**HC**, (n = 68)	**PP**, (n = 128)
Occupation		
Home duties	-	6 (4.7%)
Retired	4 (5.9%)	29 (23%)
Student	7 (10%)	7 (5.5%)
Unable to work (disability/work cover/carer)	1 (1.5%)	17 (13%)
Unemployed	1 (1.5%)	5 (3.9%)
Working - full time	39 (57%)	36 (28%)
Working - part time/casual	16 (24%)	28 (22%)
Financial status		
Finding it difficult	2 (2.9%)	16 (12%)
Getting by	23 (34%)	51 (40%)
Living comfortably	43 (63%)	61 (48%)
Place of residence		
Large town	6 (8.8%)	27 (21%)
Major city	35 (51%)	59 (46%)
Remote community	4 (5.9%)	6 (4.7%)
Small to medium town	23 (34%)	36 (28%)
Country of residence		
Australia	34 (50%)	99 (77%)
Canada	-	1 (0.8%)
Indonesia	1 (1.5%)	-
United Kingdom of Great Britain and Northern Ireland	30 (44%)	27 (21%)
United States of America	3 (4.4%)	1 (0.8%)
Country of birth		
Australia	27 (40%)	76 (59%)
Cambodia	-	1 (0.8%)
Canada	1 (1.5%)	1 (0.8%)
China	1 (1.5%)	-
Germany	-	1 (0.8%)
India	1 (1.5%)	-
Indonesia	1 (1.5%)	-
Ireland	-	1 (0.8%)

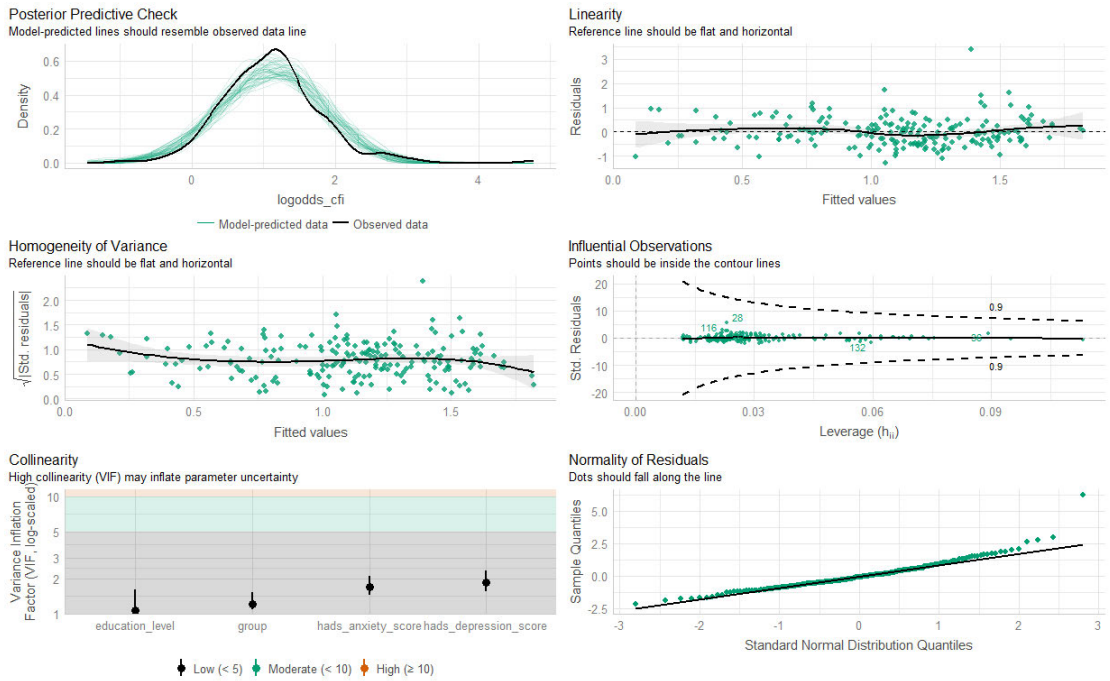
Characteristic	**HC**, (n = 68)	**PP**, (n = 128)
Italy	-	1 (0.8%)
Malaysia	-	1 (0.8%)
New Zealand	1 (1.5%)	5 (3.9%)
Pakistan	-	1 (0.8%)
South Africa	-	1 (0.8%)
The Netherlands	-	1 (0.8%)
The Philippines	1 (1.5%)	1 (0.8%)
The United States of America	3 (4.4%)	4 (3.1%)
United Kingdom of Great Britain and Northern Ireland	32 (47%)	33 (26%)
Refugee status		
No	-	127 (100%)
Missing	-	1
Race/ethnicity		
No	56 (82%)	113 (88%)
Yes	12 (18%)	15 (12%)
Neurological condition		
None	66 (97%)	122 (95%)
Stroke	2 (2.9%)	5 (3.9%)
Traumatic brain injury	-	1 (0.8%)
Mental health comorbidities		
>2	6 (8.8%)	25 (20%)
1	13 (19%)	39 (30%)
None	49 (72%)	64 (50%)
Mental health condition		
ADHD	-	5 (4%)
Alcohol Use Disorder	-	1 (1%)
Anxiety Disorders	9 (13%)	30 (23%)
Autism Spectrum Disorder	-	2 (2%)
Bipolar Disorders	1 (1%)	1 (1%)
Complex PTSD	-	6 (5%)
Depressive Disorders	14 (21%)	42 (33%)
Dissociative Identity Disorder	-	1 (1%)
Feeding and Eating Disorders	-	1 (1%)
Gender Dysphoria	1 (1%)	-
Misophonia	-	1 (1%)

Characteristic	**HC**, (n = 68)	**PP**, (n = 128)
OCD	-	2 (2%)
Personality Disorders	1 (1%)	2 (2%)
PTSD	4 (6%)	13 (10%)

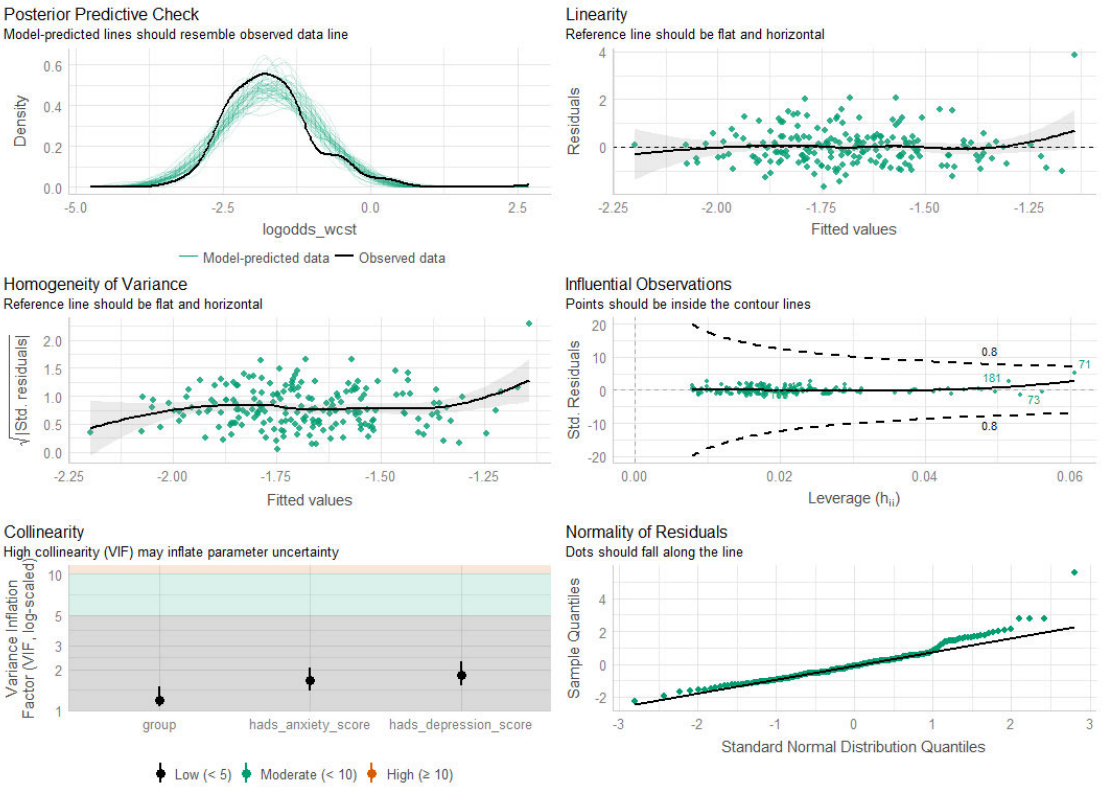
Note. ADHD = attention-deficit hyperactivity disorder; HC = healthy control; OCD = obsessive-compulsive disorder; PP = persistent pain; PTSD = post-traumatic stress disorder. Values and percentages for *mental health condition* do not add up to the total *n* for the persistent pain group (128 or 100%) because participants could report having more than one mental health condition.

Supplementary 12 (S12): Diagnostic plots for multiple linear regression models

Proportion of the Cognitive Flexibility Inventory (total score)

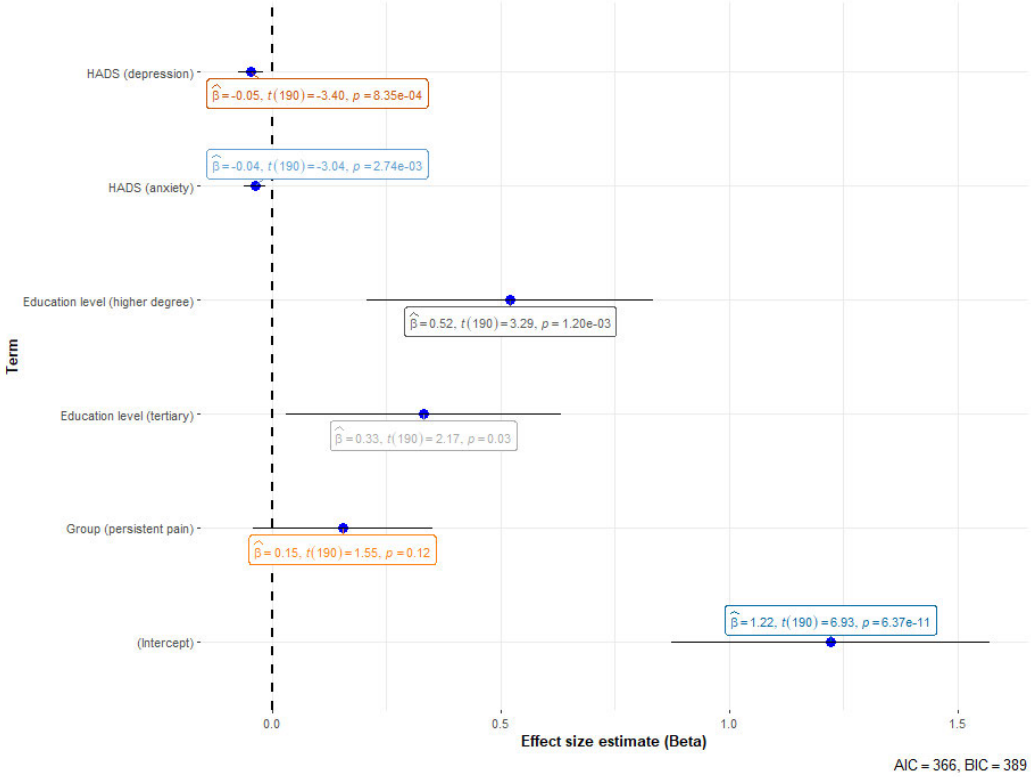


Proportion of the Wisconsin Card Sorting Test (% perseverative responses)

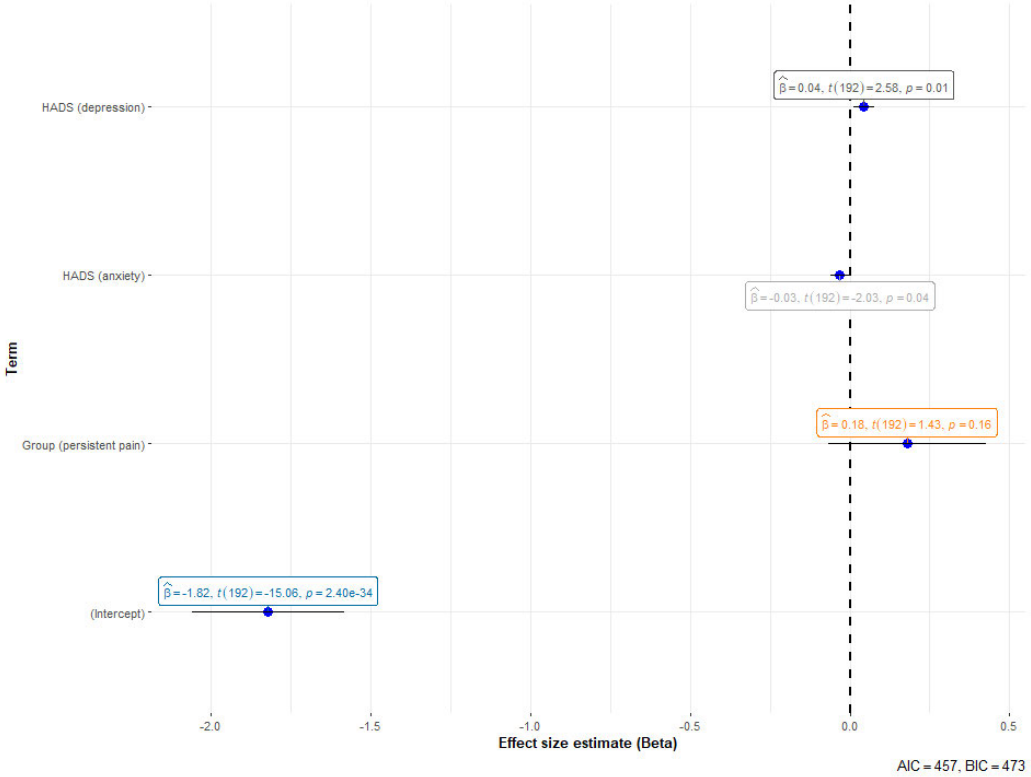


Supplementary 13 (S13): Graphical depiction of the estimates for the final model parameters

Proportion of the Cognitive Flexibility Inventory (total score)

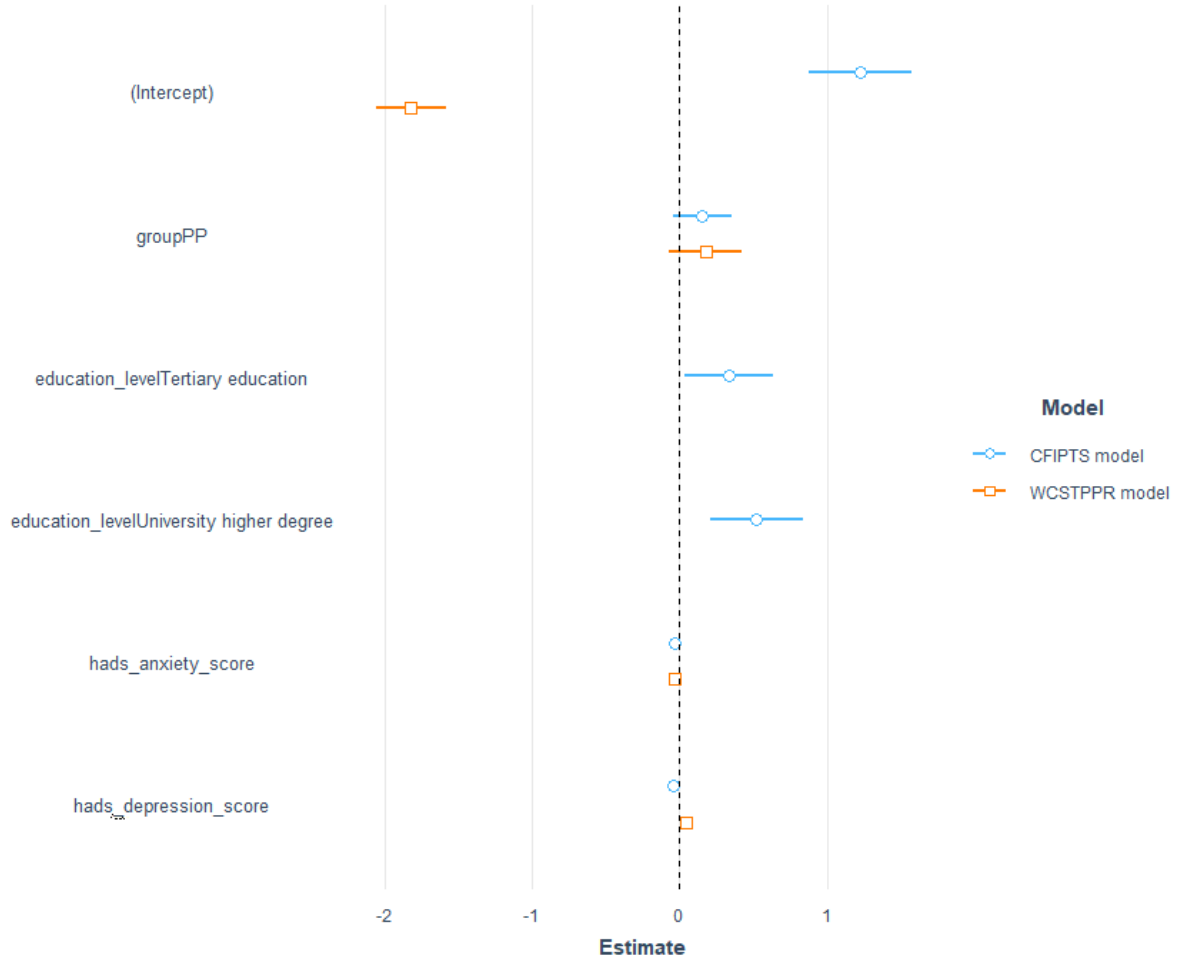


Proportion of the Wisconsin Card Sorting Test (% perseverative responses)



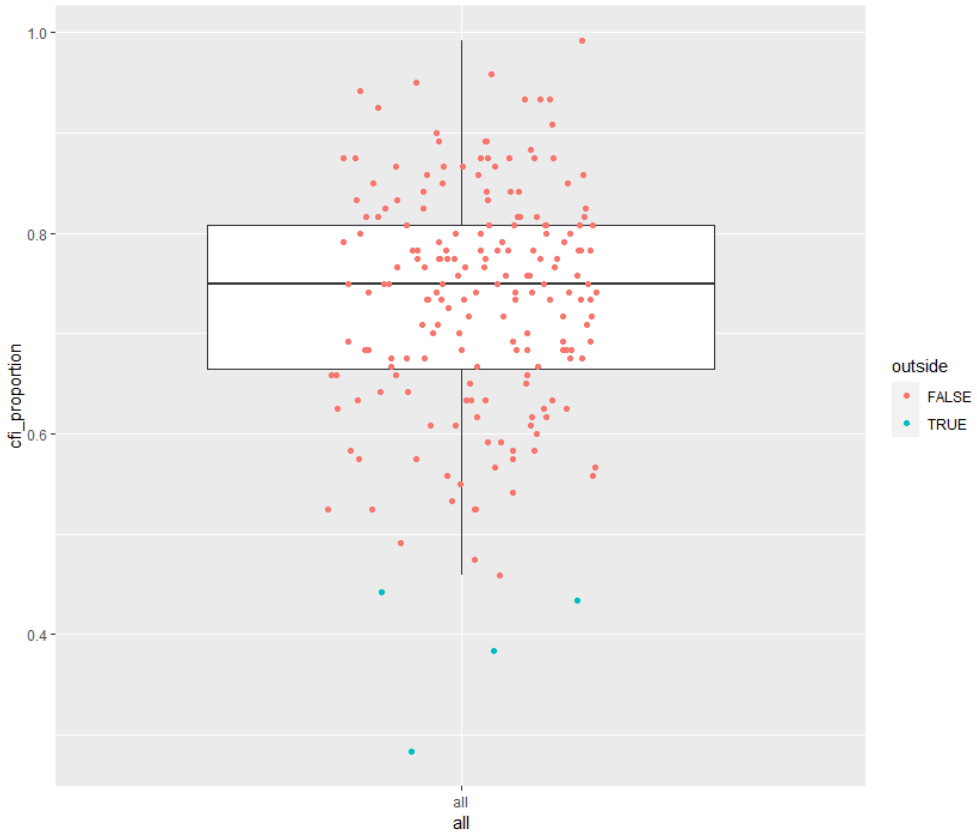
Supplementary 14 (S14): Side-by-side graphical depiction of the estimates for the final model parameters

The proportion of the Cognitive Flexibility Inventory (total score) and the proportion of the Wisconsin Card Sorting Test (% perseverative responses)

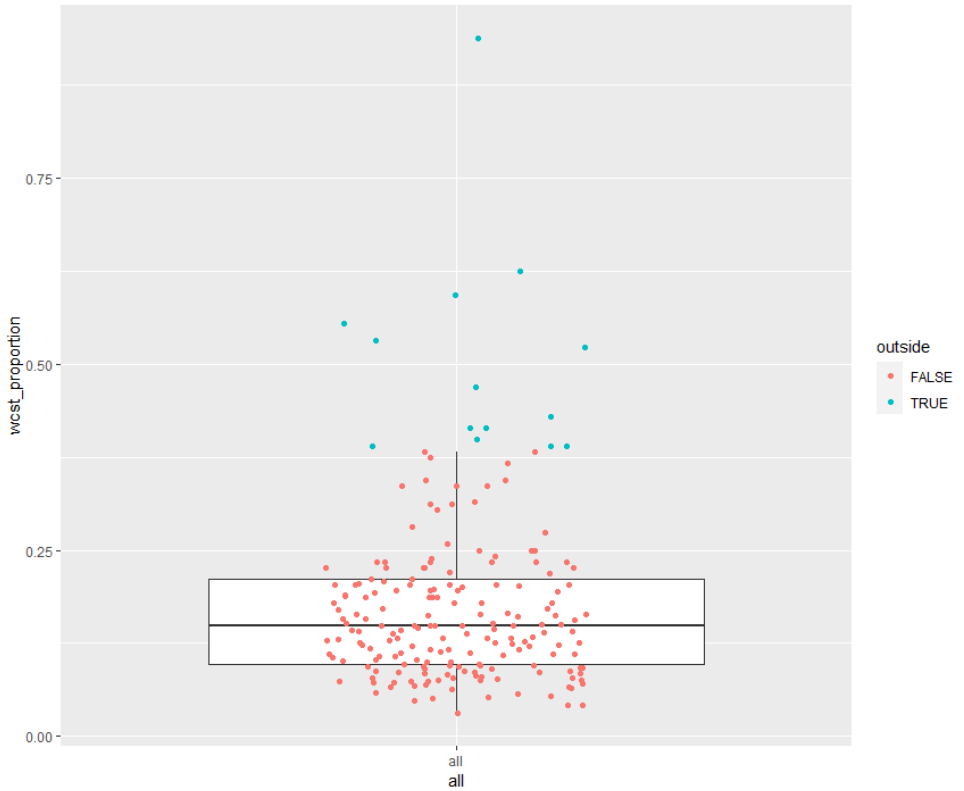


Supplementary 15 (S15): Boxplots depicting outliers for multiple linear regression models

Proportion of the Cognitive Flexibility Inventory (total score)

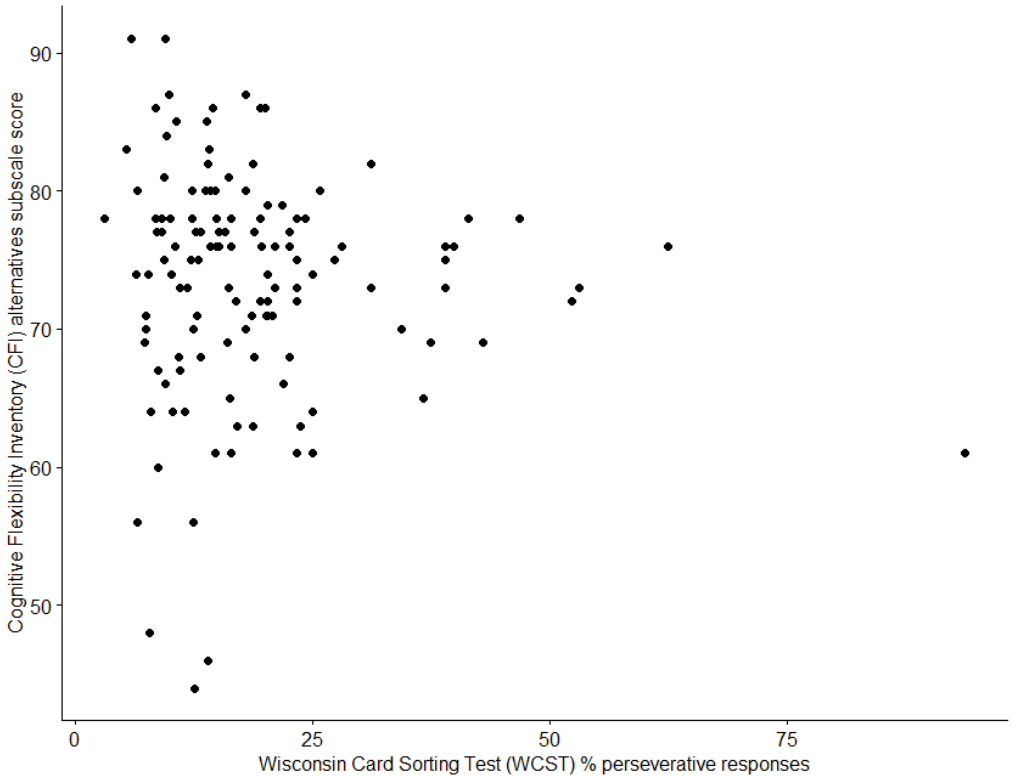


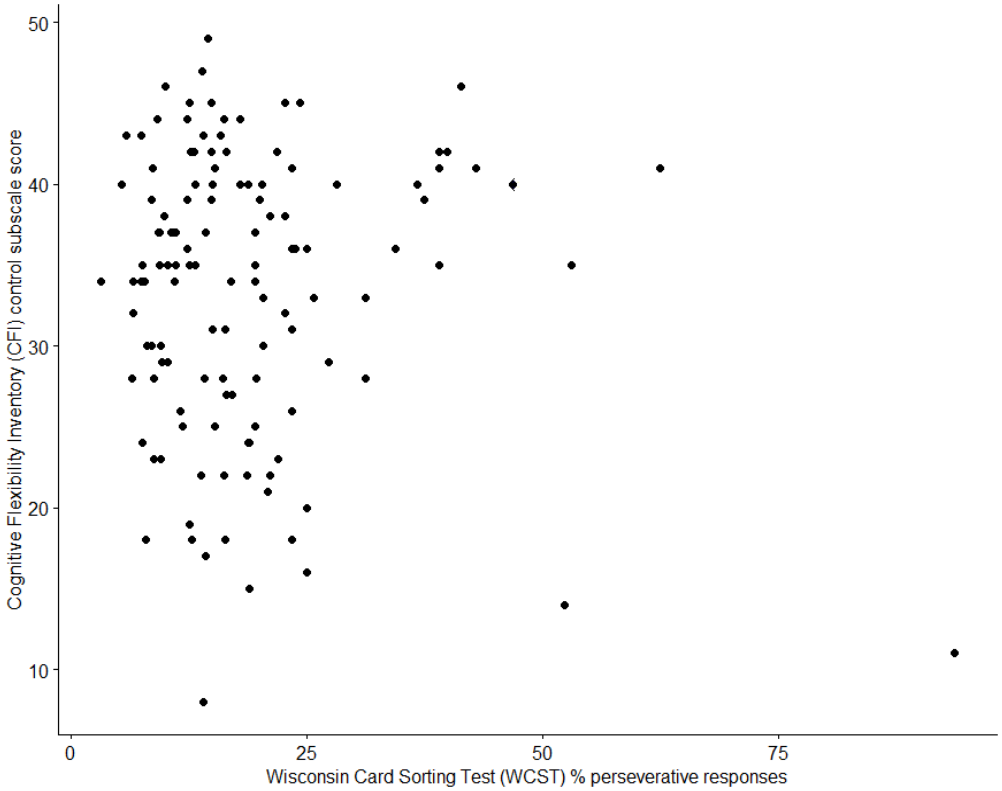
Proportion of the Wisconsin Card Sorting Test (% perseverative responses)



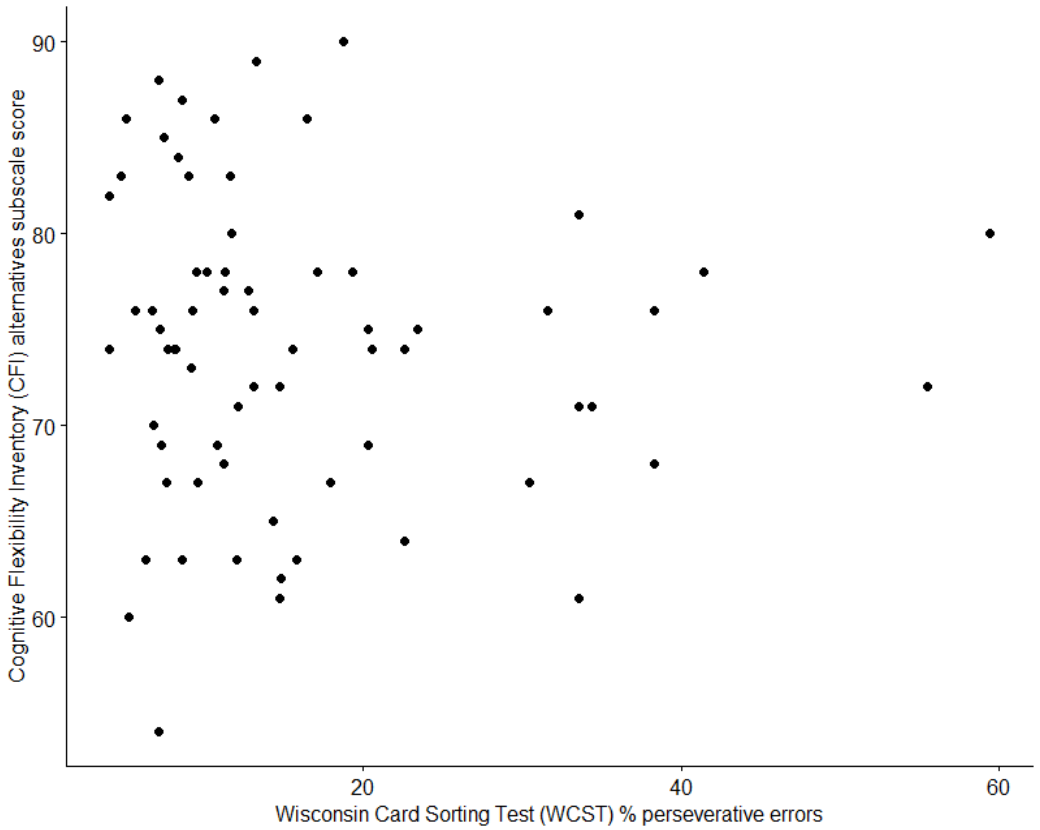
Supplementary 16 (S16): Diagnostic plots for correlation analysis

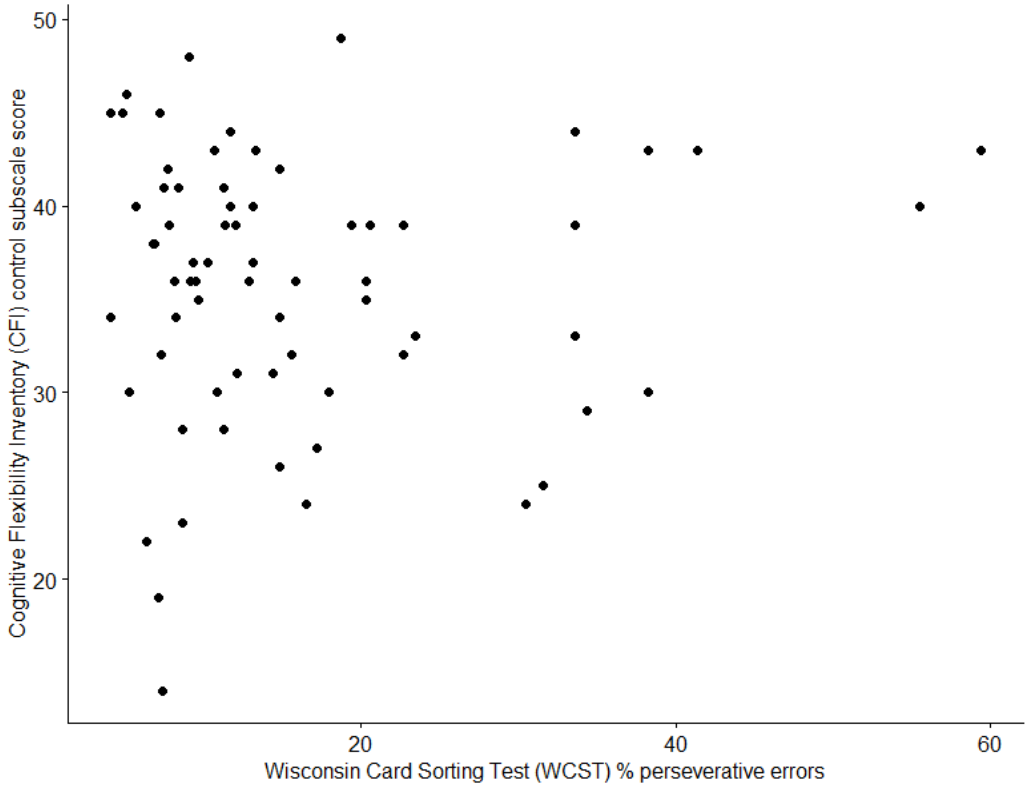
Persistent pain group





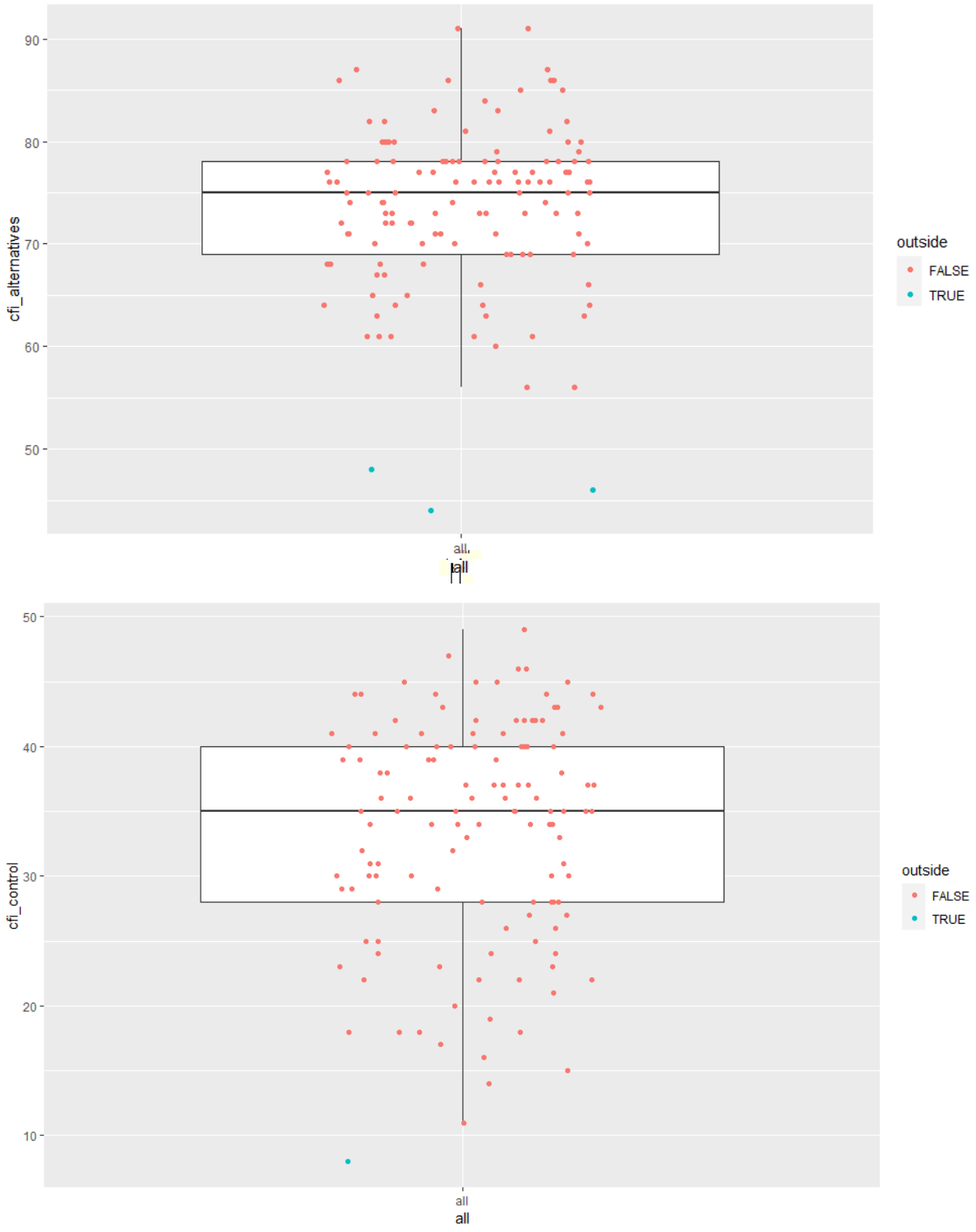
Pain-free control group

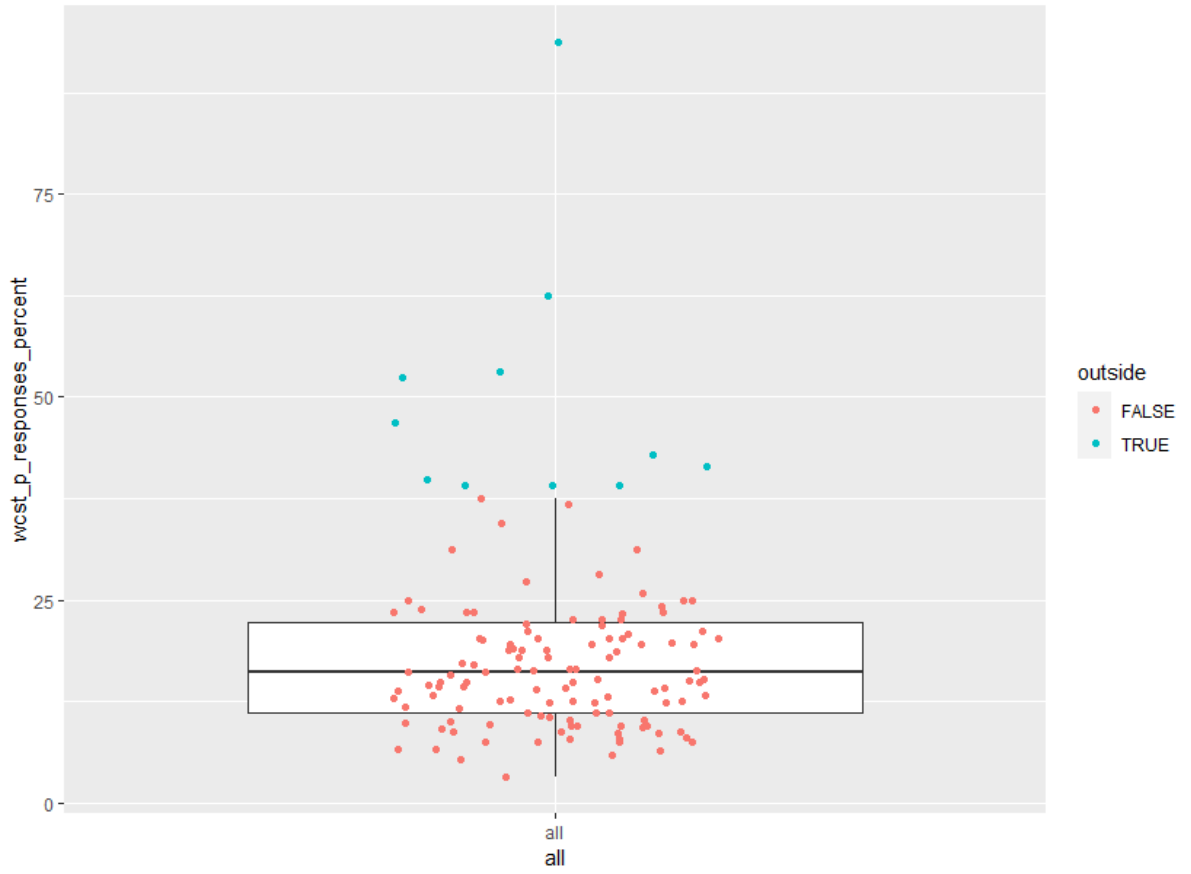




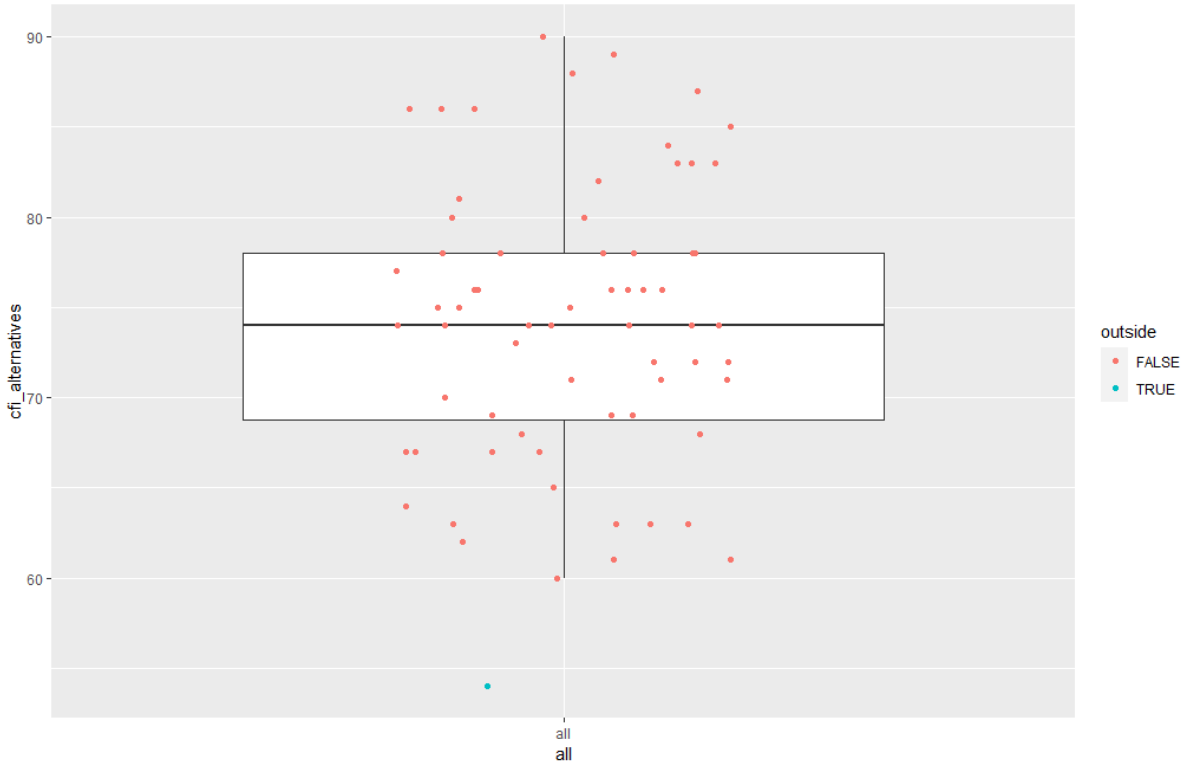
Supplementary 17 (S17): Boxplots depicting outliers for correlation analysis

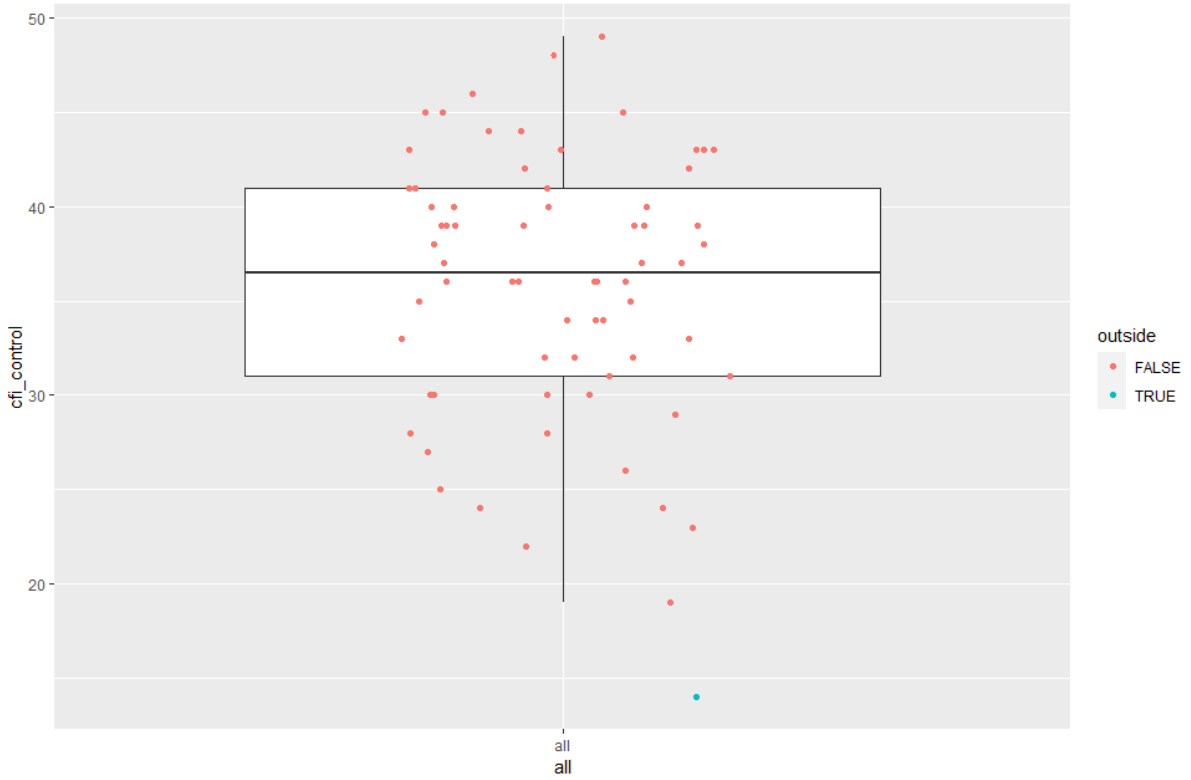
Persistent pain group

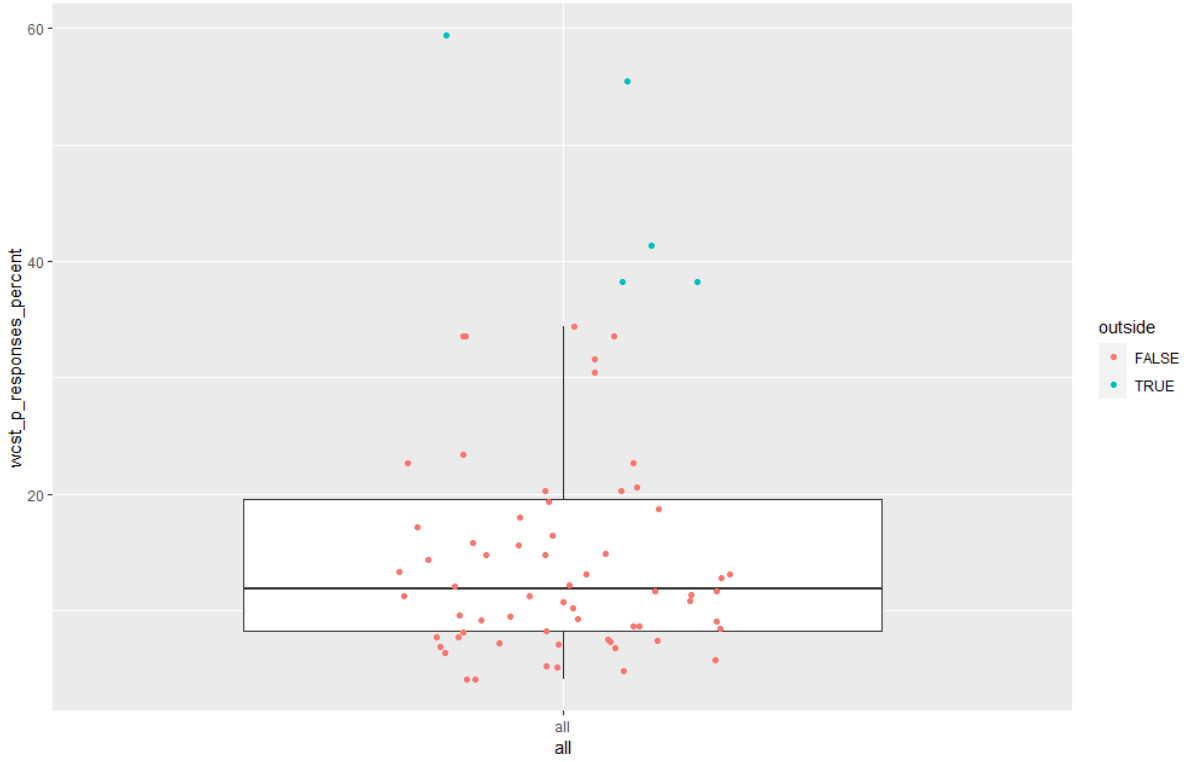




Pain-free control group

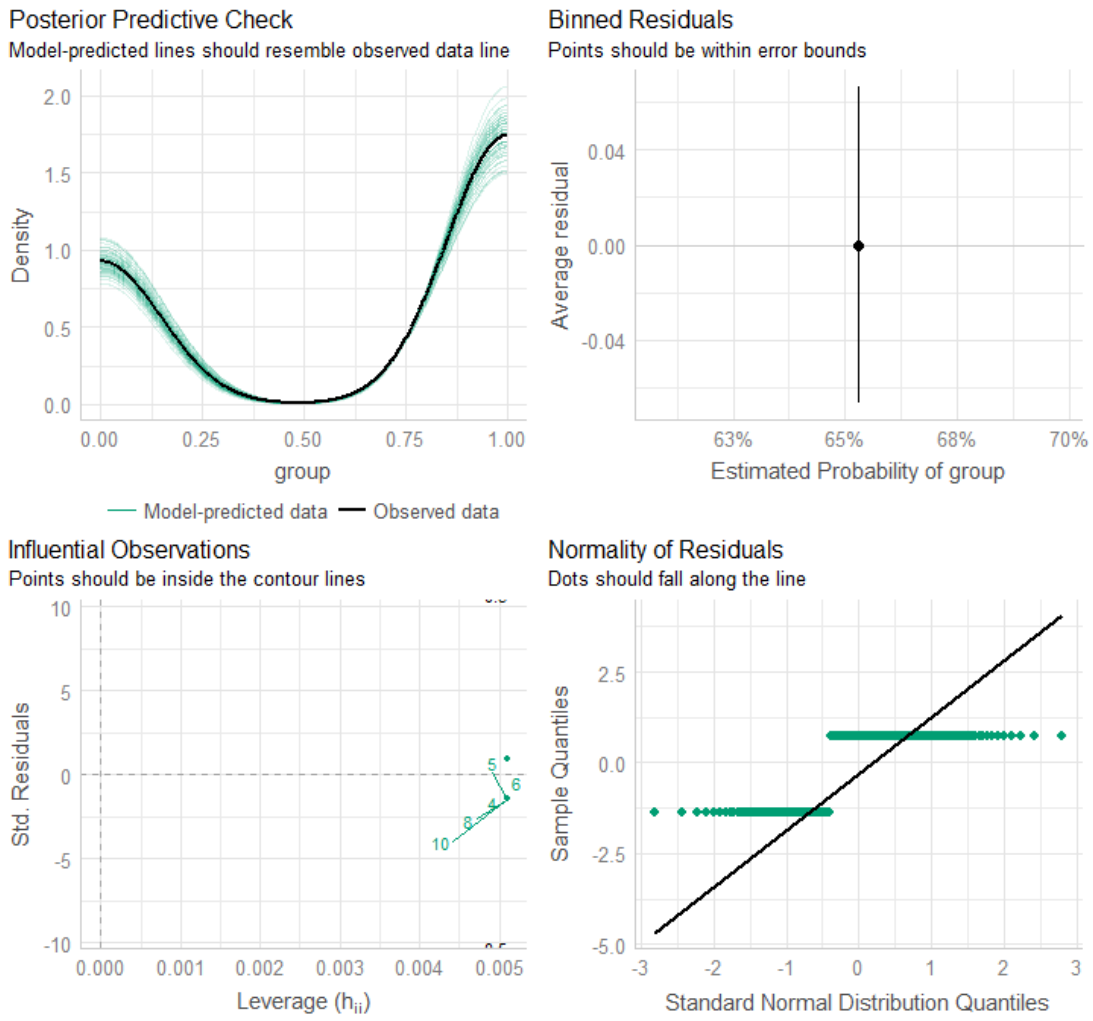



50
40
30
20
cfi_control
all
all
outside
FALSE
TRUE


60
40
20
wcst_p_responses_percent
all
all
outside
FALSE
TRUE

Supplementary 18 (S18): Diagnostic plots for logistic regression model

Note: The following diagnostic plots are redundant because the final logistic model was an intercept only model (i.e., no variance in the predictors). Additionally, the Chi-squared test for overdispersion is unnecessary for the intercept only model.



Supplementary 19 (S19): Statistical analysis plan

Statistical Analysis Plan (SAP)

Investigating Self-report and Performance-based Assessments of Cognitive Flexibility in People with and without Persistent Pain: An Online, Cross-sectional, Observational Study

Caitlin A. Howlett (IIMPACT in Health, UniSA) & Ty Stanford (ARENA, ALH, UniSA)

22 Feb 2023

Contents

1 Administrative Information **3**
1.1 Authorship Team . . . 3
1.2 Affiliations . . . 3
1.3 Ethical Approval . . . 3
1.4 Document Revision History . . . 3

2 List of Abbreviations **4**

3 Background and Rationale **5**
3.1 3.1 Synopsis . . . 5
3.2 Study Objective . . . 5
3.3 Research Questions and Hypotheses . . . 5
3.3.1 Primary Research Questions: . . . 5
3.3.2 Secondary Research Questions: . . . 5
3.3.3 Exploratory Research Question: . . . 6

4 Participants **7**
4.1 Inclusion Criteria: . . . 7
4.2 Exclusion Criteria: . . . 7
4.2.1 Participant Flowchart . . . 7

5 Measurement and Calculation of Variables **8**
5.1 Independent Variables . . . 8
5.2 Covariates . . . 8
5.3 Outcome Variables . . . 9
5.3.1 Definitions . . . 9
5.3.2 WCST . . . 9
5.3.3 CFI . . . 9

6 Data Dictionary **10**

6.1 `demographics_survey.csv` . . . 10

6.2 `demographics_summary.csv` . . . 13

6.3 `cfi_survey.csv` . . . 14

6.4 `cfi_summary.csv` . . . 15

6.5 `hads_summary.csv` . . . 16

6.6 `wcst_summary.csv` . . . 17

7 Statistical Analysis **18**

7.1 Statistical Software . . . 18

7.2 Data Cleaning . . . 18

7.3 Handling of Missing Data . . . 19

7.4 Descriptive Statistics . . . 19

7.4.1 Descriptives Table . . . 20

7.5 Primary Analysis: Multiple Linear Regression . . . 23

7.5.1 Statistical Significance and Confidence Intervals . . . 25

7.6 Secondary Analysis: Correlational Analysis . . . 25

7.6.1 Statistical Significance and Confidence Intervals: . . . 25

7.7 Exploratory Analysis: Logistic Regression . . . 26

7.7.1 Statistical Significance and Confidence Intervals . . . 26

7.7.2 Post-hoc Logistic Regression Analysis: . . . 27

7.8 Sensitivity Analyses . . . 28

7.9 Data Visualisation . . . 28

7.10 Deviations from Statistical Analysis Plan . . . 28

8 Sample Size Calculation **29**

References **30**

1 Administrative Information

1.1 Authorship Team

- Caitlin A. Howlett[1]
- Ty Stanford[2]
- Carolyn Berryman[1]
- Emma L. Karran[1]
- Valeria Bellan[1]
- Scott Coussens[3]
- Stephanie Miles[4]
- G. Lorimer Moseley[1]

1.2 Affiliations

1. Innovation, Implementation & Clinical Translation (IIMPACT) in Health, University of South Australia, Kaurna Country, Adelaide, Australia

2. Alliance for Research in Exercise, Nutrition and Activity (ARENA), Allied Health & Human Performance (ALH), University of South Australia, Kaurna Country, Adelaide, Australia

3. Cognitive and Systems Neuroscience Research Hub (CSNRH), Kaurna Country, Adelaide, Australia

4. Swinburne University of Technology, Wurundjeri Country, Melbourne, Australia

1.3 Ethical Approval

This study has been approved by UniSA's Human Research Ethics Committee (protocol no. 202439).

1.4 Document Revision History

Version	Date	Revision description	Author
1.0	21 Jul 2022	Original version	Caitlin A. Howlett and Ty Stanford
1.1	10 Aug 2022	Updated version	Caitlin A. Howlett and Ty Stanford
1.2	22 Feb 2023	Updated version	Caitlin A. Howlett and Ty Stanford

2 List of Abbreviations

Abbreviation	Meaning
CFI	Cognitive Flexibility Inventory
CFITS	*CFI total score*
CFIPTS	*CFI proportion of the total score*
HADS	Hospital Anxiety and Depression Scale
HC	Healthy control group (non-PP group)
logit	log-odds function, $\text{logit}(p) = ln(\frac{p}{1-p})$
PP	Persistent pain group
WCST	Wisconsin Card Sorting Test
WCST#PR	*WCST number of perseverative responses*
WCST%PR	*WCST percentage of perseverative responses*
WCSTPPR	*WCST proportion of perseverative responses*

3 Background and Rationale

3.1 3.1 Synopsis

This study is an online, cross-sectional, observational study looking at how people with and without persistent pain differ on two assessments of cognitive flexibility. The protocol of this study can be accessed via Open Science Framework: https://osf.io/9q8ut/. This statistical analysis plan was written after data collection, but before the commencement of data analysis.

3.2 Study Objective

To investigate how people with persistent pain perform on neurocognitive and self-reported cognitive flexibility assessments, relative to healthy controls.

3.3 Research Questions and Hypotheses

3.3.1 Primary Research Questions:

1. Do people with persistent pain perform worse than healthy controls on a neurocognitive assessment of cognitive flexibility? This question is being addressed using the following statistical hypotheses:
    - H_0: People with persistent pain perform equally to that of healthy controls on a neurocognitive assessment of cognitive flexibility.
    - H_a: People with persistent pain perform worse than healthy controls on a neurocognitive assessment of cognitive flexibility.

Note: It is expected that the alternative hypothesis will be true.

2. Is self-reported cognitive flexibility lower in people with persistent pain than it is in healthy controls? This question is being addressed using the following statistical hypotheses:
    - H_0: People with persistent pain report equal self-reported cognitive flexibility to that of healthy controls.
    - H_a: People with persistent pain report lower self-reported cognitive flexibility than healthy controls.

Note: It is expected that the alternative hypothesis will be true.

3.3.2 Secondary Research Questions:

1. Is there a relationship between a self-report and neurocognitive test of cognitive flexibility in people with persistent pain? (If so, what is the strength and direction of the association?) This question is being addressed using the following statistical hypotheses:
    - H_0: A self-report and neurocognitive assessment of cognitive flexibility are not correlated (no linear relationship) in the sub-group: people with persistent pain.
    - H_a: A self-report and neurocognitive assessment of cognitive flexibility have a non-zero correlation coefficient (linear relationship) in the sub-group: people with persistent pain.

Note: It is expected that the null hypothesis will be true.

2. Is there a relationship between a self-report and neurocognitive test of cognitive flexibility in healthy controls? (If so, what is the strength and direction of the association?) This question is being addressed using the following statistical hypotheses:

    - H_0: A self-report and neurocognitive assessment of cognitive flexibility are not correlated (no linear relationship) in the sub-group: healthy controls.

    - H_a: A self-report and neurocognitive assessment of cognitive flexibility have a non-zero correlation coefficient (linear relationship) in the sub-group: healthy controls.

Note: It is expected that the null hypothesis will be true.

3.3.3 Exploratory Research Question:

1. Is the probability of having persistent pain associated with a person's scores on a self-report and neurocognitive assessment of cognitive flexibility?

We hypothesise that there are five possibilities roughly increasing in order of complexity:

- The probability of having persistent pain (or not) is *not* associated with either a *self-reported assessment of cognitive flexibility* and a *neurocognitive assessment of cognitive flexibility.*

- The probability of having persistent pain (or not) is *not* associated with a *self-reported assessment of cognitive flexibility* but is associated with a *neurocognitive assessment of cognitive flexibility.*

- The probability of having persistent pain (or not) is associated with a *self-reported assessment of cognitive flexibility* but is *not* associated with a *neurocognitive assessment of cognitive flexibility.*

- The probability of having persistent pain (or not) is associated with a *self-reported assessment of cognitive flexibility* and a *neurocognitive assessment of cognitive flexibility* independently (on the probability of having persistent pain) (that is, the effect of a *self-reported assessment of cognitive flexibility* does not influence the effect of a *neurocognitive assessment of cognitive flexibility*, or vice-versa)

- The probability of having persistent pain (or not) is associated with a *self-reported assessment of cognitive flexibility* and a *neurocognitive assessment of cognitive flexibility*, *AND* these effects are moderated by one another (that is, the effect of a *self-reported assessment of cognitive flexibility* on the outcome of the probability of having persistent pain is dependent on the value of a *neurocognitive assessment of cognitive flexibility*, and vice-versa)

For further details regarding the exploratory research question, refer to section 7.7.

4 Participants

4.1 Inclusion Criteria:

- $\geq$ 18 years of age
- Living with or without persistent pain (i.e., pain experienced most days for three months or longer)
- Ability to read and understand English
- Access to, and competency with, a computer/laptop or smartphone with internet access
- Normal or corrected to normal vision (e.g., glasses or contact lenses)

4.2 Exclusion Criteria:

- Colourblindness

4.2.1 Participant Flowchart

Participant samples sizes will be reported in accordance with the STROBE guidelines (Vandenbroucke et al. 2014), as is shown in Figure 1 below.

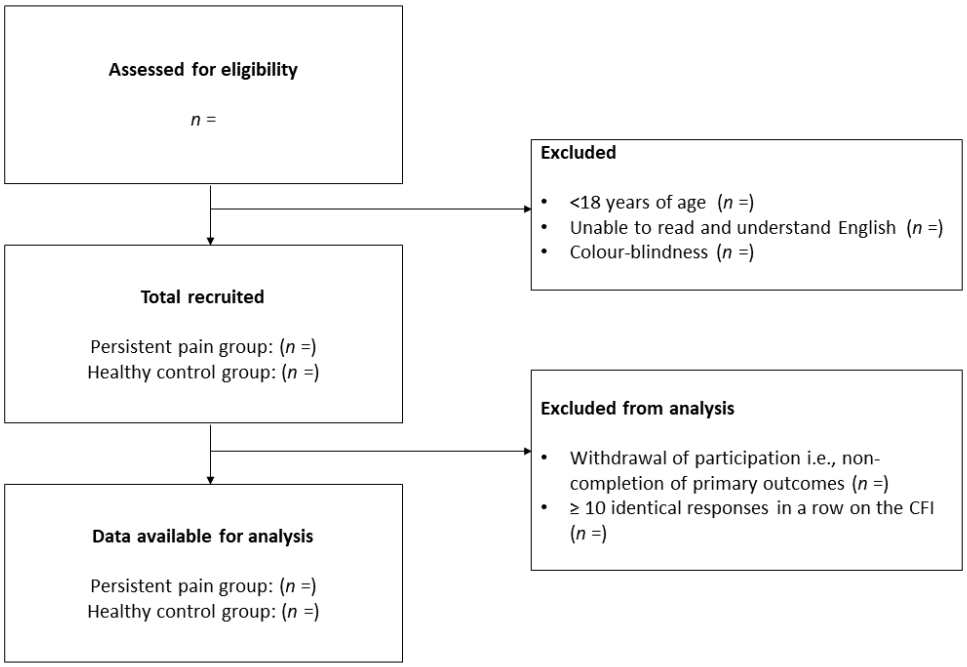


Figure 1: Flow diagram

5 Measurement and Calculation of Variables

5.1 Independent Variables

- Group: Persistent pain (PP) or healthy controls (HC)

5.2 Covariates

- Test administration order (completion of the Cognitive Flexibility Inventory (CFI) or the Wisconsin Card Sorting Test (WCST) first) `[dichotomous variable]`
- Survey completion method (modality: computer or smartphone) `[dichotomous variable]`
- Age `[continuous variable]`
- Gender (response options: male, female, trans male, trans woman, agender, non-binary) `[nominal categorical variable]`
- Education level (response options: primary school, secondary school (Year 12 or equivalent), tertiary education (trade qualification/certificate/diploma/bachelor's degree), university higher degree (masters, phd)) `[ordinal categorical variable]`
- Medication use (response options: yes or no) `[dichotomous variable]`
- Hospital Anxiety and Depression Scale (HADS) anxiety score `[treated as a continuous variable]`
- HADS depression score `[treated as a continuous variable]`

A brief review of the existing literature has revealed the following suggestions in accounting for different gender identities in research:

- "If sample sizes did not permit analysis at the level of non-binary identities, or participants would have been identifiable if this information had been shared, this should be highlighted for the benefit of future research" (Frohard-Dourlent et al. 2017)
- "If researchers choose to erase the experiences of nonbinary and trans participants, this should be clearly and distinctly reported, as well as justified (for example, with specific concerns about identifiability)" (Gofman et al. 2021)
- "If there are any 'trolling' responses, those should be removed and their existence noted in the discussion (Gofman et al. 2021)
- "Like ethnicity data, gender identity data could be dummy coded for use in regression analyses" (Fraser 2018)
- "If data from transgender participants is too limited to allow statistical analyses, descriptive statistics can, and should, be reported in a table" (Fraser 2018)
- "Researchers should also take care not to use gender as a covariate (which often results in transgender participants being excluded from analysis) without theoretical rationale. If gender must be included researchers have the option of running analyses with and without gender as a covariate, then reporting whether this changed the results" (Fraser 2018)

It should also be noted that the language and terminology used here is based on current understandings of gender identity, and it remains possible that language and terminology will evolve and change over time. With the aforementioned points in mind, this is how we plan to account for gender in our analyses:

If there are sufficient data for each of the observed gender identities (i.e., >5 participants), then each gender identify will be included in the descriptive statistics and the primary analyses as a covariate (with a possible sensitivity analysis). If data from any of the observed gender identities is too small to permit descriptive and inferential analyses (i.e., <5 participants), appropriate decisions (e.g., re-coding or collapsing of gender

identities, or exclusion of particular gender identities) will be made to ensure, on the one hand, that gender diverse people are represented in our research, and that the privacy and safety of the research participants is protected on the other. All decisions and processes used in relation to cleaning, transforming or excluding gender identities will be documented openly and transparently. Where necessary, the Trans Language Primer [n.d] will be used to guide and justify any decisions made during this process as recommended by Gofman et al. (2021). If the authorship team are not familiar with any reported gender, a basic Internet search or the following resource (www.gender.wikia.com) will be used to ensure that the responses are genuine (Fraser 2018). Any observations that are deemed to be satirical genders or 'trolling responses' (i.e., self-entered responses that are either aggressive or mischievous in nature) will be excluded from the analyses, reported as such in the participant flowchart and addressed in the discussion section of the final manuscript (Fraser 2018; Gofman et al. 2021).

5.3 Outcome Variables

5.3.1 Definitions

5.3.2 WCST

The WCST *number of perseverative responses* (WCST#PR) variable is the sum of the total perseverative responses of a participant in a maximum of 128 trials. The WCST *percent of perseverative responses* (WCST%PR) variable is therefore the WCST#PR divided by the total number of trials for a participant and can therefore be expressed as:

$$100\left(\frac{\text{WCST\#PR}}{\text{\#trials undertaken in test}}\right)$$

In line with the Heaton et al. (1993) manual, Miles et al. (2021) define *perseverative responses* as:

- "when the dimension the participant is sorting by matches the perseverated-to-principle, i.e., the participant continually responds using an incorrect dimension even after receiving feedback that this response is incorrect. Participants cannot perseverate on the *other* dimension".

Therefore, the WCST#PR for a given participant is a random variable taking the values $Y_{\#WCST} \in [0, 127]$ and WCST%PR is a random variable taking values $Y_{\%WCST} = 100 \times Y_{\#WCST} / \{\#\text{ trials undertaken in test}\} \in [0, 99.2]$ (percent). Note that values above 99.2% are not possible as the perseverative responses have to be one less than the number of the total trials undertaken (or less) by definition.

Equivalently, as the outcome variable is bounded at minimum and maximum values, we define WCST *proportion of perseverative responses* (WCSTPPR) as $Y_{p,WCST} = Y_{\#WCST} / \{\#\text{ trials undertaken in test}\} \in [0, 0.992]$.

5.3.3 CFI

The CFI is a self-report questionnaire with 20 questions in total. The CFI total score (CFITS) is the sum of the 20 Likert-scale {1, 2, 3, 4, 5, 6, 7} responses (some items are reverse scored as outlined by Dennis and Vander Wal (2010)). Therefore, the values of the CFITS random variable range from $Y_{CFITS} \in [20 - 140]$. Equivalently, as the outcome variable is bounded at minimum and maximum values, we define CFI *proportion of the total score* (CFIPTS) as $Y_{p,CFI} = (Y - 20)/120 \in [0, 1]$.

6 Data Dictionary

The variables collected in the survey and their expected output from the Inquisit software are listed in the tables below. Note that different parts of the survey are output as their own files but can be linked per participant by their subject ID.

6.1 `demographics_survey.csv`

Column	Variable Type	Comments	Expected Values
subject	alphanumeric	Respondent identifier	6 digit alphanumeric code
group	numeric	Test administration order allocation (CFI or WCST first)	1 or 2
age_eligibility_response	character	Are you at least 18 years of age?	Yes or No
colourblind_eligibility_response	character	Are you colour-blind?	Yes or No
English_eligibility_response	character	Can you read and understand English?	Yes or No
email_question1_response	character	Do you wish to receive a summary of the results for this study?	Yes or No
email_question2_response	character	Would you like to be contacted about future research?	Yes or No
email_response	character	If you ticked **yes** to either question above, please leave your email address in the text-box below:	Text response
age_response	number	What is your age? Please state your age in years _____	$\geq$ 18 years
sex_response	character	What sex were you assigned at birth?	Male, Female, Intersex or Not listed here
gender_response	character	What is your current gender identity?	Male, Female, Trans male, Trans woman, Agender or Non-binary
genderother_response	character	My gender identity is not listed. I identify as: _____	Text response

Column	Variable Type	Comments	Expected Values
education_level_response	character	What is your highest form of education?	Primary school, Secondary school, Tertiary education or University higher degree
occupation_response	character	How would you best describe your current main activity/occupation?	Unemployed, Working - full time, Working - part time/casual, Unable to work (disability/work cover/carer), Student, Home duties or Retired
financial_response	character	Which of the following statements best describes your current financial situation?	Finding it difficult, Getting by or Living comfortably
city_rural_response	character	How would you best describe where you live?	Major city, Large town, Small to medium town or Remote community
country_live_response	character	In what country do you live? _____	Text response
country_birth_response	character	What was your country of birth? _____	Text response
refugee_response	character	Do you identify as being a refugee?	Yes or No
race_ethnicity_response	character	Do you identify with a particular ethnic or racial group?	Yes or No
race_ethnicityother_response	character	If yes, please state what group you identify with: _____	Text response
mental_health_response	character	Have you previously been diagnosed with a mental health condition?	Yes or No
mental_healthother_response	character	If yes, please state all that apply: _____	Text response

Column	Variable Type	Comments	Expected Values
neurologicaloption1_response	character	Have you previously been diagnosed with any of the following conditions?	Stroke
neurologicaloption2_response	character	Have you previously been diagnosed with any of the following conditions?	Traumatic brain injury
neurologicaloption3_response	character	Have you previously been diagnosed with any of the following conditions?	Dementia or Alzheimer's disease
neurologicaloption4_response	character	Have you previously been diagnosed with any of the following conditions?	None of the above
medications_response	character	In the past week have you taken any of the following medications: Opioids, Antidepressants or Gabapentin (e.g., Lyrica)?	Yes or No
chronic_pain_question_response	character	According to this definition, do you currently have chronic pain?	Yes or No
pain_intensity_response	numeric	On a scale from 0 to 10, where **0 = no pain at all and 10 = worse pain imaginable**, how would you rate your average pain in the past week?	0 to 10
pain_duration_response	number	How long have you been experiencing chronic pain? (Please state the number of months)	≥ 3
pain_location1_response	character	Please select your **worst** area of pain	Back, Head, Neck, Abdominal, Hip, Pelvic, Hand or arm, Foot or leg or I find it difficult to identify my worst area of pain
pain_location2option1_response	character	Where else do you have pain? Please select all that apply:	Back
pain_location2option2_response	character	Where else do you have pain? Please select all that apply:	Head

Column	Variable Type	Comments	Expected Values
pain_location2option3_response	character	Where else do you have pain? Please select all that apply:	Neck
pain_location2option4_response	character	Where else do you have pain? Please select all that apply:	Abdominal
pain_location2option5_response	character	Where else do you have pain? Please select all that apply:	Hip
pain_location2option6_response	character	Where else do you have pain? Please select all that apply:	Pelvic
pain_location2option7_response	character	Where else do you have pain? Please select all that apply:	Hand or arm
pain_location2option8_response	character	Where else do you have pain? Please select all that apply:	Foot or leg
pain_location2option9_response	character	Where else do you have pain? Please select all that apply:	I do not have pain anywhere else in my body
pain_diagnosis_response	character	Have you received a diagnosis for your chronic pain?	Yes or No
pain_diagnosisother_response	character	If yes, please specify: _____	Text response

6.2 `demographics_summary.csv`

Column	Variable Type	Comments	Expected Values
subjectid	alphanumeric	Respondent identifier	6 digit alphanumeric code
groupid	numeric	Test administration order allocation (CFI or WCST first)	1 or 2
completed	numeric	0 = script was not completed (prematurely aborted); 1 = script was completed (all conditions run)	0 or 1
computer.platform	character	Survey completion method	win, mac, ios or and

6.3 `cfi_survey.csv`

Column	Variable Type	Comments	Expected Values
subject	alphanumeric	Respondent identifier	6 digit alphanumeric code
group	numeric	Test administration order allocation (CFI or WCST first)	1 or 2
q1_response	numeric	I am good at "sizing up" situations	Strongly disagree = 1 to Strongly agree = 7
q2R_response	numeric	I have a hard time making decisions when faced with difficult situations	Strongly disagree = 1 to Strongly agree = 7
q3_response	numeric	I consider multiple options before making a decision	Strongly disagree = 1 to Strongly agree = 7
q4R_response	numeric	When I encounter difficult situations, I feel like I am losing control	Strongly disagree = 1 to Strongly agree = 7
q5_response	numeric	I like to look at difficult situations from many different angles	Strongly disagree = 1 to Strongly agree = 7
q6_response	numeric	I seek additional information not immediately available before attributing causes to behavior	Strongly disagree = 1 to Strongly agree = 7
q7R_response	numeric	When encountering difficult situations, I become so stressed that I can not think of a way to resolve the situation	Strongly disagree = 1 to Strongly agree = 7
q8_response	numeric	I try to think about things from another person's point of view	Strongly disagree = 1 to Strongly agree = 7
q9R_response	numeric	I find it troublesome that there are so many different ways to deal with difficult situations	Strongly disagree = 1 to Strongly agree = 7
q10_response	numeric	I am good at putting myself in others' shoes	Strongly disagree = 1 to Strongly agree = 7
q11R_response	numeric	When I encounter difficult situations, I just don't know what to do	Strongly disagree = 1 to Strongly agree = 7
q12_response	numeric	It is important to look at difficult situations from many angles	Strongly disagree = 1 to Strongly agree = 7
q13_response	numeric	When in difficult situations, I consider multiple options before deciding how to behave	Strongly disagree = 1 to Strongly agree = 7

Column	Variable Type	Comments	Expected Values
q14_response	numeric	I often look at a situation from different viewpoints	Strongly disagree = 1 to Strongly agree = 7
q15_response	numeric	I am capable of overcoming the difficulties in life that I face	Strongly disagree = 1 to Strongly agree = 7
q16_response	numeric	I consider all the available facts and information when attributing causes to behavior	Strongly disagree = 1 to Strongly agree = 7
q17R_response	numeric	I feel I have no power to change things in difficult situations	Strongly disagree = 1 to Strongly agree = 7
q18_response	numeric	When I encounter difficult situations, I stop and try to think of several ways to resolve it	Strongly disagree = 1 to Strongly agree = 7
q19_response	numeric	I can think of more than one way to resolve a difficult situation I'm confronted with	Strongly disagree = 1 to Strongly agree = 7
q20_response	numeric	I consider multiple options before responding to difficult situations	Strongly disagree = 1 to Strongly agree = 7

Note: R = reverse scored

6.4 `cfi_summary.csv`

Column	Variable Type	Comments	Expected Values
subjectid	alphanumeric	Respondent identifier	6 digit alphanumeric code
groupid	numeric	Test administration order allocation (CFI or WCST first)	1 or 2
completed	numeric	0 = script was not completed (prematurely aborted); 1 = script was completed (all conditions run)	0 or 1
computer.platform	character	Survey completion method	win, mac, ios or and
cfi_alternatives	numeric	Alternative subscale score	13 to 91
cfi_control	numeric	Control subscale score	7 to 49
cfi_total	numeric	Total score	20 to 140

6.5 `hads_summary.csv`

Column	Variable Type	Comments	Expected Values
subjectid	alphanumeric	Respondent identifier	6 digit alphanumeric code
groupid	numeric	Test administration order allocation (CFI or WCST first)	1 or 2
completed	numeric	0 = script was not completed (prematurely aborted); 1 = script was completed (all conditions run)	0 or 1
computer.platform	character	Survey completion method	win, mac, ios or and
HADS_DepressionScore	numeric	Total score	0 to 21
HADS_AnxietyScore	numeric	Total score	0 to 21
radiobuttons.q1R_A.response	numeric	I feel tense or 'wound up'	0 to 3
radiobuttons.q2_D.response	numeric	I still enjoy the things I used to enjoy	0 to 3
radiobuttons.q3R_A.response	numeric	I get a sort of frightened feeling as if something awful is about to happen	0 to 3
radiobuttons.q4_D.response	numeric	I can laugh and see the funny side of things	0 to 3
radiobuttons.q5R_A.response	numeric	Worrying thoughts go through my mind	0 to 3
radiobuttons.q6R_D.response	numeric	I feel cheerful	0 to 3
radiobuttons.q7_A.response	numeric	I can sit at ease and feel relaxed	0 to 3
radiobuttons.q8R_D.response	numeric	I feel as if I am slowed down	0 to 3
radiobuttons.q9_A.response	numeric	I get a sort of frightened feeling like 'butterflies' in the stomach	0 to 3
radiobuttons.q10R_D.response	numeric	I have lost interest in my appearance	0 to 3
radiobuttons.q11R_A.response	numeric	I feel restless as I have to be on the move	0 to 3

Column	Variable Type	Comments	Expected Values
radiobuttons.q12_D.response	numeric	I look forward with enjoyment to things	0 to 3
radiobuttons.q13R_A.response	numeric	I get sudden feelings of panic	0 to 3
radiobuttons.q14_D.response	numeric	I can enjoy a good book or radio or TV program	0 to 3

Note: R = reverse scored; D = depression scale; A = anxiety scale

6.6 `wcst_summary.csv`

Column	Variable Type	Comments	Expected Values
subjectid	alphanumeric	Respondent identifier	6 digit alphanumeric code
groupid	numeric	Test administration order allocation (CFI or WCST first)	1 or 2
completed	numeric	0 = script was not completed (prematurely aborted); 1 = script was completed (all conditions run)	0 or 1
computer.platform	character	Survey completion method	win, mac, ios or and
number_pResponses	numeric	Perseverative responses (sum) score	0 to 127
Percent_pResponses	numeric	Perseverative responses (percent) score	0% to 99.2%
number_pErrors	numeric	Perseverative errors (sum) score	0 to 127
Percent_pErrors	numeric	Perseverative errors (percent) score	0% to 99.2%

Note: Although other variables may be included in the output of the raw data files, they are not listed here for conciseness. With the exception of questions relating to identifiable information (i.e., participant email addresses), only the variables listed in the data dictionary will be used for descriptive or (to assist with) inferential analyses.

7 Statistical Analysis

7.1 Statistical Software

All statistical analyses will be performed using the current stable version of R (R Core Team 2022) and the `sessionInfo()` will be reported.

7.2 Data Cleaning

Prior to data analysis, identifying information (i.e., participant email addresses) will be removed and stored separately from the study data in a password protected file. Data wrangling will be performed using the `tidyverse` (Wickham et al. 2019) and `janitor` (Firke 2023) packages. Separate data files will be exported from Inquisit Web. Using the `dplyr` (Wickham et al. 2022) package, separate data files will then be merged into a single file by subject ID. Data collected prior to participant withdrawal (non-completion of primary outcomes) will not be analysed as per UniSA's Human Research Ethics Committee protocol (no. 202439). Participant data will be excluded from the analyses if 10 or more identical responses are provided in a row on the CFI. No data will be excluded based on participants' performance on the WCST. Where duplicate participants (i.e., same user ID) are present in the data set, the first entry will be retained for analysis and any subsequent entries will be excluded. See Figure 2 for full details regarding the data cleaning process.

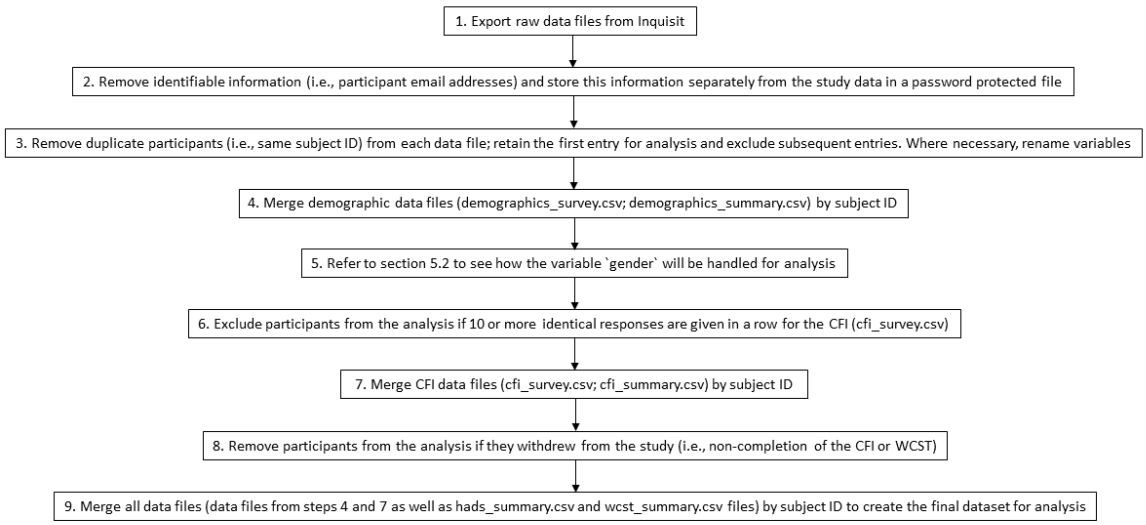


Figure 2: Data cleaning flow diagram

7.3 Handling of Missing Data

If more than 5% of the participants contained in the final analysis records have one or more missing values, primary analyses will be performed with multiple imputation using a fully conditional specification approach (i.e., multiple imputation by chained equations) irrespective of the missingness mechanism (i.e., missing completely at random (MCAR), missing at random (MAR), or not missing at random (NMAR)). Multiple imputation will be performed using the `mice` package (Van Buuran and Groothuis-Oudshoorn 2011), which uses a multivariate imputation by chained equations algorithm. Prior to undertaking any statistical analyses, numeric imputed values will be checked for anomalies e.g., a) ensure that the imputed values do not exceed the pre-defined bounds of the observed values of the variables, and in turn, produce nonsensical values, and b) ensure that the imputed values do not show substantially different patterns in scatterplots/histograms when compared to the observed values. If anomalies are present, the imputation model will be adjusted such that a predictive mean matching (PMM) approach will be used for numeric variables in lieu of a linear regression approach. Built-in default imputation methods will be used for incomplete binary and categorical variables. Refer to Table 1 in Van Buuran and Groothuis-Oudshoorn (2011) for further information. Little's (1988) MCAR test will be used to test the MCAR assumption across the entire original data set (observed values and missing values). If Little's test is statistically significant ($p<0.05$) the null hypothesis that there is no difference between the observed and missing mean values will be rejected, i.e., there is a difference between the observed and missing mean values, implying the missing data are not MCAR. Alternatively, if Little's test is not statistically significant ($p>0.05$), the null hypothesis that there is no difference between the observed and missing mean values will be retained. The number of imputed data sets will be set to 50 to minimise simulation error (Van Buuran and Groothuis-Oudshoorn 2011). Multiple imputation will be performed irrespective of (a) the missingness mechanism or (b) the outcome of Little's MCAR test as outlined by (Van Ginkel et al. 2020). However, if Little's test is statistically significant, this will be noted in the results so that interpretation can be made with this in mind. All variables used in the primary analyses will be used to perform multiple imputation as outlined by Van Buuran and Groothuis-Oudshoorn (2011) because variables related to non-response will not be able to be identified. If multiple imputation is used, the pooled analysis will be used to determine statistical significance and the model estimates (regression coefficient) with standard errors, which will be combined using Rubin's (1987) rule. Using Rubin's rule to combine separate estimates and standard errors from each imputed dataset is appropriate when using regression coefficients (White, Royston, and Wood 2011). Sterne et al. (2009) and the STROBE guidelines (Vandenbroucke et al. 2014) will be used to guide the reporting of the final manuscript.

7.4 Descriptive Statistics

Descriptive statistics will be reported in a summary table using raw values/scores, similar to what is depicted in the example below (see section 7.4.1). Continuous variables will be presented as means and standard deviations if the data are normally distributed, or as medians and interquartile ranges if the data are skewed. The range of values for continuous variables will be reported. Dichotomous and categorical data will be presented as numbers and percentages for each level. All descriptive statistics will be reported for each group separately. The number and percentage of missing values for each variable will also be reported in the descriptives table.

7.4.1 Descriptives Table

Variables	Persistent Pain (n = x)	Controls (n = x)
Age (in years)	x (x)	x (x)
Missing	x (%)	x (%)
Range	x-x	x-x
Sex		
Male	x (%)	x (%)
Female	x (%)	x (%)
Intersex	x (%)	x (%)
Not listed	x (%)	x (%)
Missing	x (%)	x (%)
Gender		
Male	x (%)	x (%)
Female	x (%)	x (%)
Trans male	x (%)	x (%)
Trans woman	x (%)	x (%)
Agender	x (%)	x (%)
Non-binary	x (%)	x (%)
xx	x (%)	x (%)
xx	x (%)	x (%)
Missing	x (%)	x (%)
Education level		
Primary school	x (%)	x (%)
Secondary school	x (%)	x (%)
Tertiary education	x (%)	x (%)
University higher degree	x (%)	x (%)
Missing	x (%)	x (%)
Occupation		
Unemployed	x (%)	x (%)
Working (full time)	x (%)	x (%)
Working (part time/casual)	x (%)	x (%)
Unable to work (disability/work cover/carer)	x (%)	x (%)
Student	x (%)	x (%)
Home duties	x (%)	x (%)
Retired	x (%)	x (%)
Missing	x (%)	x (%)
Financial		
Finding it difficult	x (%)	x (%)
Getting by	x (%)	x (%)
Living comfortably	x (%)	x (%)
Missing	x (%)	x (%)
Place of residence		
Major city	x (%)	x (%)
Large town	x (%)	x (%)
Small to medium town	x (%)	x (%)
Remote community	x (%)	x (%)

Variables	Persistent Pain (n = x)	Controls (n = x)
Missing	x (%)	x (%)
Country of residence		
xx	x (%)	x (%)
xx	x (%)	x (%)
xx	x (%)	x (%)
Missing	x (%)	x (%)
Country of birth		
xx	x (%)	x (%)
xx	x (%)	x (%)
xx	x (%)	x (%)
Missing	x (%)	x (%)
Refugee		
Yes	x (%)	x (%)
No	x (%)	x (%)
Missing	x (%)	x (%)
Racial/Ethnic group		
Yes	x (%)	x (%)
No	x (%)	x (%)
xx	x (%)	x (%)
xx	x (%)	x (%)
Missing	x (%)	x (%)
Mental health condition		
No	x (%)	x (%)
Yes	x (%)	x (%)
xx	x (%)	x (%)
xx	x (%)	x (%)
Neurological condition		
Stroke	x (%)	x (%)
Traumatic brain injury	x (%)	x (%)
Dementia or Alzheimer's disease	x (%)	x (%)
None of the above	x (%)	x (%)
Medications (Opioids, Antidepressants or Gabapentin)		
Yes	x (%)	x (%)
No	x (%)	x (%)
Pain intensity	x (x)	x (x)
Range	x-x	x-x
Pain duration (in months)	x (x)	x (x)
Range	x-x	x-x
Pain location (primary)		
Back	x (%)	x (%)
Neck	x (%)	x (%)

Variables	Persistent Pain (n = x)	Controls (n = x)
Hip	x (%)	x (%)
Abdominal	x (%)	x (%)
Pelvic	x (%)	x (%)
Hand or arm	x (%)	x (%)
Foot or leg	x (%)	x (%)
Pain location (secondary)		
Back	x (%)	x (%)
Neck	x (%)	x (%)
Hip	x (%)	x (%)
Abdominal	x (%)	x (%)
Pelvic	x (%)	x (%)
Hand or arm	x (%)	x (%)
Foot or leg	x (%)	x (%)
Diagnosis		
No	x (%)	x (%)
Yes	x (%)	x (%)
xx	x (%)	x (%)
xx	x (%)	x (%)
Cognitive Flexibility Inventory		
CFI alternatives subscale	x (x)	x (x)
Range	x-x	x-x
CFI control subscale	x (x)	x (x)
Range	x-x	x-x
CFI total score	x (x)	x (x)
Range	x-x	x-x
Wisconsin Card Sorting Test		
WCST perseverative responses (sum)	x (x)	x (x)
Range	x-x	x-x
WCST perseverative responses (percent)	x (x)	x (x)
Range	x-x	x-x
WCST perseverative errors (sum)	x (x)	x (x)
Range	x-x	x-x
WCST perseverative errors (percent)	x (x)	x (x)
Range	x-x	x-x
Hospital Anxiety and Depression Scale		
HADS-Anxiety	x (x)	x (x)
Range	x-x	x-x
HADS-Depression	x (x)	x (x)
Range	x-x	x-x

Note: x (x) = central tendency and dispersion values; x (%) = number and percentage values; xx = unspecified responses; x-x = range of values.

7.5 Primary Analysis: Multiple Linear Regression

Backwards stepwise model selection will be performed to create a parsimonious regression model for both the primary outcomes. Of primary interest is to investigate the effect of group (PP and HC) on the outcome variables (and will thus remain in the final model). As outlined in sections 5.3.2 and 5.3.3, the outcome variables will be calculated as proportions. To ensure that the linear model assumptions of unbounded outcomes hold, the outcome variables will be transformed using log-odds transformation, $\text{logit}(p) = ln(\frac{p}{1-p})$. In the unlikely case of proportions equal to 0 (or 1), a small constant will be added (subtracted) to avoid undefined logit transformed values. The constants used will be a half increment of the possible values in the WCST and CFI proportion scores, that is, $\frac{1}{2\times 127}$ and $\frac{1}{2\times 120}$, respectively (i.e., $\frac{1}{2\times \text{maximum value}}$).

Separate models will be constructed for each outcome variable.

The log-odds transformed linear model can be expressed in the general form:

$$logit(Y_i) = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i$$

Where:

- $i = 1, 2, ..., n$ is the participant number,
- Y_i is the outcome variable (a proportion) for the i^{th} participant,
- β_0 is an intercept (the predicted value of $logit(Y_i)$ when all predictors are 0),
- p is the number of predictors,
- X_{ij} is the j^{th} predictor for the i^{th} participant,
- β_j is the coefficient of the j^{th} predictor, and
- $\epsilon_i \sim N(0, \sigma)$ (independently and identically distributed normal error)

However, the full model for WCSTPPR takes the following specific form:

$$\begin{aligned}
&\text{logit}(Y_{i,WCSTPPR}) = \beta_0 + \beta_1 \times I(group_i = PP) + \beta_2 \times age_i + \\
&\beta_3 \times I(education_i = secondary\ school) + \\
&\beta_4 \times I(education_i = tertiary) + \\
&\beta_5 \times I(education_i = university\ higher\ degree) + \\
&\beta_6 \times I(gender_i = female) + \\
&\beta_7 \times I(gender_i = trans\ male) + \\
&\beta_8 \times I(gender_i = trans\ woman) + \\
&\beta_9 \times I(gender_i = agender) + \\
&\beta_{10} \times I(gender_i = non-binary) + \\
&\beta_{11} \times I(medication_i = yes) + \\
&\beta_{12} \times HADSanxiety_i + \\
&\beta_{13} \times HADSdepression_i + \\
&\beta_{14} \times I(test\ order_i = group\ 2) + \\
&\beta_{15} \times I(modality_i = smartphone) + \epsilon_i
\end{aligned}$$

Where:

- $I(.)$ is the indicator function. That is, $I(x) = 1$ if x is true and $I(x) = 0$ if x is false.

Similarly, the full model for CFIPTS takes the following specific form:

$$
\begin{aligned}
&\text{logit}(Y_{i,CFIPTS}) = \beta_0 + \beta_1 \times I(group_i = PP) + \beta_2 \times age_i + \\
&\beta_3 \times I(education_i = secondary\ school) + \\
&\beta_4 \times I(education_i = tertiary) + \\
&\beta_5 \times I(education_i = university\ higher\ degree) + \\
&\beta_6 \times I(gender_i = female) + \\
&\beta_7 \times I(gender_i = trans\ male) + \\
&\beta_8 \times I(gender_i = trans\ woman) + \\
&\beta_9 \times I(gender_i = agender) + \\
&\beta_{10} \times I(gender_i = non - binary) + \\
&\beta_{11} \times I(medication_i = yes) + \\
&\beta_{12} \times HADSanxiety_i + \\
&\beta_{13} \times HADSdepression_i + \\
&\beta_{14} \times I(test\ order_i = group\ 2) + \\
&\beta_{15} \times I(modality_i = smartphone) + \epsilon_i
\end{aligned}
$$

Where:

- $I(.)$ is the indicator function. That is, $I(x) = 1$ if x is true and $I(x) = 0$ if x is false.

The assumptions of linearity, homoscedasticity of the residuals variance, and influential observations will be visually inspected using scatterplots; normality of residuals will be inspected via a Q-Q plot, and standardised residuals will be checked to not have associated cook's distance >=1. Multicollinearity will be checked via a bar chart. All diagnostic plots will be generated using the `performance` package (Lüdecke et al. 2021). If these assumptions are violated, alternative models and transformations will be considered. Akaike Information Criterion (AIC) will be used for model comparisons and to assess the best-fit model (i.e., goodness of fit).

7.5.1 Statistical Significance and Confidence Intervals

In the model reduction process (backwards selection) model terms will be removed in a stepwise manner where variables have p-values > 0.05, removing terms with the largest p-values first. Statistical significance will be determined at the $\alpha = 0.05$ level for our estimated group (PP vs HC) effect on the primary outcomes (WCSTPPR and CFIPTS) and will be reported with 95% $(100 \times (1 - \alpha))$ confidence intervals. Additionally, the remaining variables in the final models resulting from the backwards selection and their associated p-values, will be reported in addition to their model coefficient estimates.

For example, if `education` is selected in the final model for WCSTPPR, the corresponding reporting will take the form:

variable	p-value	coefficient	estimate
education	0.0XX	primary school	*reference*
		secondary school	X.XX
		tertiary education	X.XX
		university higher degree	X.XX

7.6 Secondary Analysis: Correlational Analysis

Pearson's r correlation coefficients and/or Spearman's rho correlation coefficients will be used to determine the strength of the relationship between variables for the secondary research questions. The assumptions of linearity and homoscedascity will be assessed through visual inspection of diagnostic plots (i.e., scatterplots). For variable pairs that meet these assumptions, a Pearson's r correlation coefficient will be calculated; where the assumptions of linearity and homoscedascity are violated, a Spearman's rho correlation coefficient will be calculated in lieu of Pearson's r. Linearity and homoscedascity will be assessed independently by three reviewers; consensus will be established when 2/3 reviewers reach agreement. Effect sizes will be interpreted as small (0.10-0.29), medium (0.30-0.49), and large (>0.50), as outlined by Cohen (1992).

Variables of interest:

CFI variables	WCST variables
Alternative subscale	Perseverative responses (%)
Control subscale	

7.6.1 Statistical Significance and Confidence Intervals:

Sidak corrected α-levels will be applied to adjust for multiple comparisons to avoid identifying false positives for the secondary research questions. Sidak adjustments to the level of significance is based on the sorted

p-values labelled $p^{(1)}, p^{(2)}, ..., p^{(n)}$ for n hypothesis tests, then each $p^{(m)}$ value ($m = 1, 2, ..., n$) is compared to its corresponding Sidak adjusted level of significance. Sidak corrected α-levels will be calculated in the form of:

$$\alpha_{Sidak}^{(m)} = 1 - (1 - \alpha)^{1/m}$$

Where:

- α is the unadjusted threshold for significance (0.05)
- m is 1 through to n based on the ordered p-values
- n is the number of tests performed ($n = 4$ here: 2 for each group)

Accordingly, adjusted $100(1 - \alpha_{Sidak}^{(m)})\%$ confidence intervals of the estimated r or rho coefficients will be reported.

7.7 Exploratory Analysis: Logistic Regression

A logistic regression will be performed to investigate the effect of two assessments of cognitive flexibility (CFI or WCST) on the probability of detecting a binary outcome (PP or HC). The predictor variables $CFIPTS$ and $WCSTPPR$ are continuous variables on the [0, 1] range. Additionally, a scatterplot will be used to visualise the relationships between the three variables. Diagnostic statistics and plots (i. Chi-square test for overdispersion, and ii. scatter plot of standardized Pearson's residuals against fitted values on the logit scale for homoskedasticity) of the logistic regression model will be generated using the `performance` package (Lüdecke et al. 2021). If these assumptions of the model fit are violated, alternative models and transformations will be considered.

The logistic regression model can be expressed as:

$$logit(p_i) = \beta_0 + \beta_1 \times CFIPTS_i + \beta_2 \times WCSTPPR_i + \beta_3 \times CFIPTS_i \times WCSTPPR_i$$
$$Y_i \sim Bernoulli(p_i)$$

Where:

- p_i is the probability of person i having persistent pain, that is, $p_i = P(group_i = \text{persistent pain})$ where $group_i$ is the group (PP or HC) of person i and $P(x)$ is the probability of event x,
- β_0 is an intercept (the predicted value of $logit(Y_i)$ when all predictors are 0),
- $CFIPTS_i$ is the *CFI proportion of the total score* for the i^{th} participant,
- $WCSTPPR_i$ is the *WCST proportion of perseverative responses (%)* for the i^{th} participant,
- Y_i is the observed PP status of person i, that is, $Y_i = 1$ if person i is in the group PP and $Y_i = 0$ if person i is not in the group PP (i.e., HC)

Note: All predictors will be centred and scaled based on sampled distributions of values (and coefficients interpreted accordingly).

7.7.1 Statistical Significance and Confidence Intervals

Statistical significance will be determined at the $\alpha = 0.05$ level for our estimated effect of the predictors (WCSTPPR and CFIPTS) on the probability of detecting our binary outcome (PP vs HC) and will be reported with 95% ($100 \times (1 - \alpha)$) confidence intervals. The backwards model selection process is outlined in Figure 3.

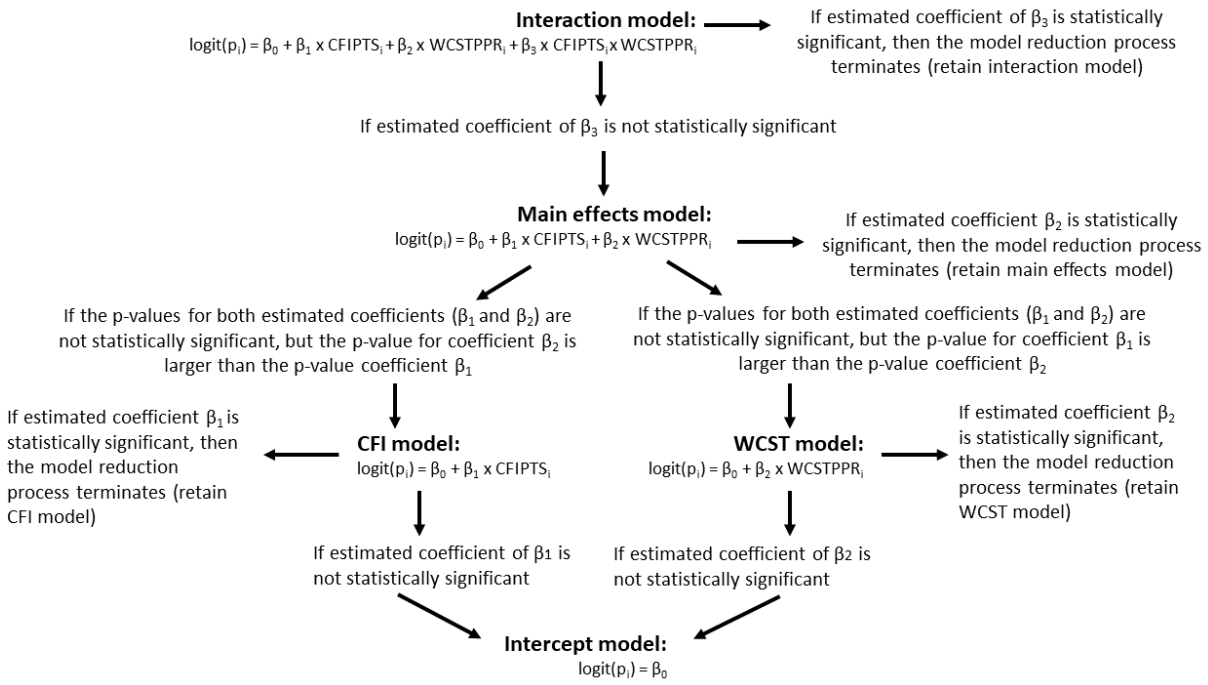


Figure 3: Logistic regression model backwards selection process

7.7.2 Post-hoc Logistic Regression Analysis:

The final logistic regression model will be used to investigate the model estimated probability of having persistent pain based on 'high' and 'low' values of the independent variables (CFIPTS and WCSTPPR, provided they remain in the final model). All combinations of 'high' and 'low' values of the independent variables will produce up to four predicted probabilities (with $100(1-\alpha)\%$ confidence intervals) as per the table in the *hypothesised subgroups* subsection. To determine values for 'high' and 'low' scores on each predictor variable, the 1st and 3rd (weighted) quartiles on the sampled data for each independent variable will be used. To account for unequal group sizes, the prevalence of persistent pain in the general adult population in both developed (Blyth et al. 2001; Breivik et al. 2006; Pitcher et al. 2019) and developing (Sá et al. 2019) countries (i.e., 1 in 5 or 20%) will inform the weights for calculating the interquartile range.

Hypothesised subgroups:

Note that the use of 'high' and 'low' values for the WCST and CFI in the post-hoc estimation from the logistic regression model assess the following hypothesised relationships in the table below:

PP Group	HC Group
Low performance on WCST, low performance on CFI	High performance on WCST, high performance on CFI
High performance on WCST, low performance on CFI	
Low performance on WCST, high performance on CFI	

Where:

- Higher % perseverative responses scores on the WCST are indicative of poor (or low) performance; lower % perseverative responses scores on the WCST are indicative of better (or high) performance.

- Higher total scores on the CFI are indicative of better (or high) self-evaluation; lower total scores on the CFI are indicative of poor (or low) self-evaluation.

The above table can equivalently be specified as per below:

PP Group	HC Group
High WCST score, low CFI score	Low WCST score, high CFI score
Low WCST score, low CFI score	
High WCST score, high CFI score	

7.8 Sensitivity Analyses

- For the primary research questions, a sensitivity analysis will be performed with any univariate outliers that are detected (defined as CFIPTS or WCSTPPR values that fall below Q1 - 1.5* IQR or exceed Q3 + 1.5* IQR, which will be inspected via box-plots) included and then again with outliers removed.
- For the secondary research questions, a sensitivity analysis will be performed with any univariate outliers that are detected (defined as CFI subscale scores or WCST % PR values that fall below Q1 - 1.5* IQR or exceed Q3 + 1.5* IQR, which will be inspected via box-plots) included and then again with outliers removed.

7.9 Data Visualisation

Figures and plots will be created using `ggplot2` (Wickham 2016) and `ggpubr` (Kassambara 2020). Where possible to do so, figures and plots produced will be created using colour-blind friendly colour palettes in `base R` and in the `grDevices` (R Core Team 2022) package to enhance inclusivity.

7.10 Deviations from Statistical Analysis Plan

Any deviations from the original Statistical Analysis Plan will be recorded in a separate document and uploaded to Open Science Framework: https://osf.io/9q8ut/ and clearly described in the final manuscript.

8 Sample Size Calculation

Sample size calculations were made under the assumptions of the analysis to be undertaken on the two primary analyses to ensure sufficiently powered primary analyses. Mean and standard deviation estimates of the outcome variables (WCST#PR and CFI scores) in previous studies (Dennis and Vander Wal 2010; Greve et al. 2005; Johnco, Wuthrich, and Rapee 2014) were utilised. A detectable increase of 5% for the outcome of WCST%PR in the PP group compared to the HC group was used, and a detectable decrease of 6 (similarly, 5% of the scale) for the outcome of CFITS in the PP group compared to the HC group was used. A required sample size of 120 was estimated in conjunction with a power of 0.95 and level of significance of 0.05. The sample size calculations were performed using simulation (Arnold et al. 2011; Green and MacLeod 2016; Kontopantelis et al. 2016) of a beta distributed response of a linear model data generating process as both outcomes are approximately continuous, skewed and bounded. Sensitivity analysis of the required sample size was also performed using alternative model specifications and a range in the number of covariates and their relative effects. The empirical calculations were verified using analytical multiple linear regression power analysis methods available in G*Power 3 (Faul et al. 2007, 2009) and the `pwr` package (Champely 2020) in R version 4.1.3 (R Core Team 2022). Refer to the document titled "Sample size calculations for WCST and CFI modelling" (pain-cog-sample-size.pdf, v1.0, 2022/04/05) for further details. Sample size calculations were undertaken prior to data analysis, but not prior to data collection.

References

Arnold, B F, D R Hogan, J M Colford, and A E Hubbard. 2011. "Simulation methods to estimate design power: An overview for applied research." *BMC Medical Research Methodology* 11: 94. https://doi.org/10.1186/1471-2288-11-94.

Blyth, F M, L M March, A J M Brnabic, L R Jorm, M Williamson, and M J Cousins. 2001. "Chronic pain in Australia: A prevalence study." *Pain* 89 (2): 127–34. https://doi.org/10.1016/S0304-3959(00)00355-9.

Breivik, H, B Collett, V Ventafridda, R Cohen, and D Gallacher. 2006. "Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment." *European Journal of Pain* 10 (4): 287–333. https://doi.org/10.1016/j.ejpain.2005.06.009.

Champely, S. 2020. "pwr: Basic functions for power analysis. R package version 1.3-0." https://cran.r-project.org/package=pwr.

Cohen, J. 1992. "Quantitative methods in psychology: A power primer." *Psychological Bulletin* 112 (1): 155–59.

Dennis, John P., and Jillon S. Vander Wal. 2010. "The cognitive flexibility inventory: Instrument development and estimates of reliability and validity." *Cognitive Therapy and Research* 34 (3): 241–53. https://doi.org/10.1007/s10608-009-9276-4.

Faul, F, E Erdfelder, A Buchner, and A-G Lang. 2009. "Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses." *Behavior Research Methods* 41 (4): 1149–60. https://doi.org/10.3758/BRM.41.4.1149.

Faul, F, E Erdfelder, A-G Lang, and A Buchner. 2007. "G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences." *Behavior Research Methods* 39 (2): 175–91. https://doi.org/10.3758/BF03193146.

Firke, S. 2023. "janitor: Simple tools for examining and cleaning dirty data. R package version 2.2.0." https://cran.r-project.org/package=janitor.

Fraser, G. 2018. "Evaluating inclusive gender identity measures for use in quantitative psychological research." *Psychology and Sexuality* 9 (4): 343–57. https://doi.org/10.1080/19419899.2018.1497693.

Frohard-Dourlent, Hélène, S Dobson, B A Clark, M Doull, and E M Saewyc. 2017. "'"I would have preferred more options": Accounting for non-binary youth in health research." *Nursing Inquiry* 24 (1): e12150. https://doi.org/10.1111/nin.12150.

Gofman, A, S A Leif, H Gunderman, and N Exner. 2021. "Do I have to be an "other" to be myself? Exploring gender diversity in taxonomy, data collection, and through the research data lifecycle." *Journal of eScience Librarianship* 10 (4): e1219. https://doi.org/10.7191/jeslib.2021.1219.

Green, P, and CJ MacLeod. 2016. "SIMR: An R package for power analysis of generalized linear mixed models by simulation." *Methods in Ecology and Evolution* 7: 493–98. https://doi.org/10.1111/2041-210X.12504.

Greve, K W, T R Stickle, J M Love, K J Bianchini, and M S Stanford. 2005. "Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study." *Archives of Clinical Neuropsychology* 20: 355–64. https://doi.org/10.1016/j.acn.2004.09.004.

Heaton, Robert K., G. J. Chelune, J. L. Talley, G. G. Kay, and G Curtiss. 1993. *Wisconsin Card Sorting Test manual: Revised and expanded*. Psychological Assessment Resources.

Johnco, C, V M Wuthrich, and R M Rapee. 2014. "Reliability and validity of two self-report measures of cognitive flexibility." *Psychological Assessment* 26 (4): 1381–87. https://doi.org/10.1037/a0038009.

Kassambara, Alboukadel. 2020. "ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0." https://cran.r-project.org/package=ggpubr.

Kontopantelis, E, D A Springate, R Parisi, and D Reeves. 2016. "Simulation-based power calculations for mixed effects modeling: ipdpower in Stata." *Journal of Statistical Software* 74 (12): 1–25. https://doi.org/10.18637/jss.v074.i12.

Little, R J A. 1988. "A test of missing completely at random for multivariate data with missing values." *Journal of the American Statistical Association* 83 (404): 1198–1202. https://doi.org/10.2307/ 2290157.

Lüdecke, D, M S Ben-Shachar, I Patil, P Waggoner, and D Makowski. 2021. "performance: An R package for assessment, comparison and testing of statistical models." *The Journal of Open Source Software* 6 (60): 3139. https://doi.org/10.21105/joss.03139.

Miles, S, C A Howlett, C Berryman, M Nedeljkovic, G L Moseley, and A Phillipou. 2021. "Considerations

for using the Wisconsin Card Sorting Test to assess cognitive flexibility." *Behavior Research Methods* 53: 2083–91. https://doi.org/10.3758/s13428-021-01551-3.

Pitcher, M H, M Von Korff, M C Bushnell, and L Porter. 2019. "Prevalence and profile of high-impact chronic pain in the United States." *The Journal of Pain* 20 (2): 146–60. https://doi.org/10.1016/j.jpain.2018.07.006.

R Core Team. 2022. "R: A language and environment for statistical computing." Vienna: R Foundation for Statistical Computing. http://www.r-project.org/.

Rubin, DB. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Sá, K N, L Moreira, A F Baptista, L T Yeng, M J Teixeira, R Galhardoni, and D C de Andrade. 2019. "Prevalence of chronic pain in developing countries: Systematic review and meta-analysis." *Pain Reports* 4 (6): e779. https://doi.org/10.1097/PR9.0000000000000779.

Sterne, J A C, I R White, J B Carlin, M Spratt, P Royston, M G Kenward, A M Wood, and J R Carpenter. 2009. "Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls." *BMJ* 338. https://doi.org/10.1136/bmj.b2393.

Van Buuran, S, and K Groothuis-Oudshoorn. 2011. "mice: Multivariate Imputation by Chained Equations in R." *2Journal of Statistical Software* 45 (3): 1–67. https://doi.org/10.18637/jss.v045.i03.

Van Ginkel, J R, M Linting, R C A Rippe, and A van der Voort. 2020. "Rebutting existing misconceptions about multiple imputation as a method for handling missing data." *Journal of Personality Assessment* 102 (3): 297–308. https://doi.org/10.1080/00223891.2018.1530680.

Vandenbroucke, J P, E von Elm, D G Altman, P C Gøtzsche, C D Mulrow, S J Pocock, C Poole, J J Schlesselman, and M Egger. 2014. "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration." *International Journal of Surgery* 12: 1500–1524. https://doi.org/10.1016/j.ijsu.2014.07.014.

White, I R, P Royston, and A M Wood. 2011. "Multiple imputation using chained equations: Issues and guidance for practice." *Statistics in Medicine* 30: 377–99. https://doi.org/10.1002/sim.4067.

Wickham, H. 2016. *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, H, M Averick, J Bryan, W Chang, L D'Agostino McGowan, R François, G Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, H, R François, L Henry, and K Müller. 2022. "dplyr: A Grammar of Data Manipulation." https://cran.r-project.org/package=dplyr.

Supplementary 20 (S20): Deviations from statistical analysis plan

Deviations from SAP

Caitlin A. Howlett (IIMPACT in Health, UniSA) & Ty Stanford (ARENA, ALH, UniSA)

22 Feb 2023

SAP Revision History

Date	Description of changes	Author
10 Aug 2022	The traditional owners of the land on which Swinburne University is located was added to an authors affiliation; the table provided in section 7.5.1 was amended because the options listed for the `education` variable were incorrectly listed.	Caitlin A. Howlett and Ty Stanford
13 Oct 2022	>10 identical responses in a row on the HADS was originally meant to prompt participant exclusion. However, we decided not to exclude participants on the basis of their HADS scores because we learned that reverse scoring the responses was unnecessary due to the nature of the Likert scale of the questionnaire. That is, the Likert scale of the HADS is not consistent for every item e.g., for some items the scale ranges from 0-3, and for other items the scale ranges from 3-0, making it very unlikely that >10 identical responses in a row would occur. The ISQ was removed because we chose not to proceed with that aspect of the study.	Caitlin A. Howlett and Ty Stanford

Appendix E: Permissions

The manuscript presented in **Chapter 2** entitled: 'Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility' was published in *Behavior Research Methods.* Springer Nature permits journal authors to reproduce their article in full, or in part, for scholarly, non-commercial purposes, one of which is inclusion in a thesis or dissertation. This also includes the right to make a thesis or dissertation available in the awarding academic institutions electronic repository or database. The formal permission request for this article can be found on page 249.

The manuscript presented in **Chapter 3** entitled: 'Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults' was published in *Clinical Psychology Review.* Elsevier permits journal authors to reproduce their article in full, or in part, for scholarly, non-commercial purposes, one of which is inclusion in a thesis or dissertation. This also includes the right to make a thesis or dissertation available in the awarding academic institutions electronic repository or database. A formal permission request for this article was not required.

The manuscript presented in **Chapter 4** entitled: 'Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis' was published in *Neuropsychology* and was reproduced with permission from the American Psychological Association (APA). The formal permission request for this article can be found on page 256.

The manuscript presented in **Chapter 5** entitled: 'Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy' was published in the *Australian Journal of Psychology* under a Creative Commons Attribution-NonCommercial (CC-BY-NC) licence; therefore, no formal permission is required to reproduce it in this thesis.

The manuscript presented in **Chapter 6** entitled: 'Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study' is not published, but under review at *The British Journal of Pain*; therefore, no formal permission is required to reproduce it in this thesis.

Copyright Clearance for Chapter Two

*Miles, S., ***Howlett, C. A.,** Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, 53, 2083-2091. https://doi.org/10.3758/s13428-021-01551-3

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS
Mar 01, 2023

This Agreement between Caitlin Howlett ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5500480946112
License date	Mar 01, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Behavior Research Methods
Licensed Content Title	Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility
Licensed Content Author	Stephanie Miles et al
Licensed Content Date	Mar 22, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	50000 or greater
Author of this Springer Nature content	yes
Title	Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility
Institution name	University of South Australia
Expected presentation date	Mar 2023
Order reference number	1204
Requestor Location	Caitlin Howlett 108 North Terrace

Adelaide, 5001
Australia
Attn: Caitlin Howlett

Total 0.00 AUD

Terms and Conditions

Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

1. **Grant and Scope of License**

    1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly permitted by this License.

    2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

    3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

    4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

    5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

2. **Reservation of Rights**

   Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

3. **Restrictions on use**

   1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not limited to, cropping, adapting, and/or omitting material that affect the meaning, intention or moral rights of the author(s) are strictly prohibited.

   2. You must not use any Licensed Material as part of any design or trademark.

   3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

4. **STM Permission Guidelines**

   1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/.

   2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and conditions contained in this License.

   3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the

Rightsholder.
Contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

5. **Duration of License**

1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings
Reuse in an annual report	Lifetime of annual report
Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

6.
7. **Acknowledgement**

1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time

as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

8. **Reuse in a dissertation or thesis**

1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/) and only up to what is required by the awarding institution.

2. For theses published under an ISBN or ISSN, separate permission is required. Please contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: '*Reproduced with permission from Springer Nature*'.

9. **License Fee**

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

10. **Warranty**

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the**

copyright of another entity or third party (as credited in the published version). If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

11. **Termination and Cancellation**

1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.

12. **General**

1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United

Nations Convention on Contracts for the International Sale of Goods) or to Germany´s choice-of-law principle.

2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

Questions? For questions on Copyright Clearance Center accounts or website issues please contact springernaturesupport@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution

Other Conditions:

Version 1.4 - Dec 2022

Questions? customercare@copyright.com.

Copyright clearance for Chapter Four

Howlett, C. A., Wewege, M. A., Berryman, C., Oldach, A., Jennings, E., Moore, E., Karran, E. L., Szeto, K., Pronk, L., Miles, S., & Moseley, G. L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology, 36*(5), 347-372.

https://doi.org/10.1037/neu0000796

AMERICAN PSYCHOLOGICAL ASSOCIATION LICENSE
TERMS AND CONDITIONS
Mar 01, 2023

This Agreement between Caitlin Howlett ("You") and American Psychological Association ("American Psychological Association") consists of your license details and the terms and conditions provided by American Psychological Association and Copyright Clearance Center.

License Number	5500490621375
License date	Mar 01, 2023
Licensed Content Publisher	American Psychological Association
Licensed Content Publication	Neuropsychology
Licensed Content Title	Back to the drawing board—The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis.
Licensed copyright line	Copyright © 2022, American Psychological Association
Licensed Content Author	Howlett, Caitlin A.; Wewege, Michael A.; Berryman, Carolyn; Oldach, Annika; Jennings, Elizabeth; Moore, Emily; Karran, Emma L.; Szeto, Kimberley; Pronk, Leander; Miles, Stephanie; Moseley, G. Lorimer
Licensed Content Date	Jul 1, 2022
Licensed Content Volume	36
Licensed Content Issue	5
Type of Use	Thesis/Dissertation
Requestor type	Academic institution
Format	Print, Electronic

Portion	page
Number of pages requested	26
Page range of portion	347-372
Rights for	Main product and any product related to main product
Duration of use	life of current edition
Creation of copies for the disabled	no
With minor editing privileges	no
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
Title	Back to the drawing board – the relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts. A systematic review and meta-analysis
Institution name	University of South Australia
Expected presentation date	Mar 2023
Order reference number	1204
Portions	All pages of the published journal article (i.e., pages 347-372, including all figures and tables included in the published article), and any supplementary files that accompany the published article.
Requestor Location	Caitlin Howlett 108 North Terrace Adelaide, 5001 Australia Attn: Caitlin Howlett
Total	0.00 AUD

Terms and Conditions

RightsLink Plus Terms and Conditions
American Psychological Association

The publisher for this copyrighted material is the American Psychological Association ("APA"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by the Copyright Clearance Center, Inc. ("CCC") at the time that you opened your CCC account.

APA hereby grants you a non-exclusive license to use this material, subject to the conditions below.

1. Limited License.
   Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process; any form of republication must be completed with 12 months from the date hereof (although copies prepared before then may be distributed thereafter); and any electronic posting is limited to a period of 12 months.

2. Credit Lines and Translation Disclaimer.
   The reproduced material must include one of the following credit lines:

   For English language content: Copyright © [year] by [Copyright Owner]. Reproduced [or Adapted] with permission. [Full bibliographic citation.]

   For translated language content: This material originally appeared in English as [full bibliographic citation]. Copyright © [year] by [Copyright Owner]. Translated and reproduced [or adapted] with permission. [Copyright Owner] is not responsible for the quality or accuracy of this translation. This translation cannot be reproduced or distributed further without prior written permission.

3. Author Permission.
   You must obtain the author's (or, in case of multiple authorship, one author's) permission. A good faith effort must be made to locate the author and obtain the author's permission. Author permission is not required for the reuse of small excepts or 1 to 3 figures or 1 to 3 tables; or for course pack or electronic reserve use.

4. Geographic Rights.
   Licenses may be exercised anywhere in the world.

5. Book Edition Limit.
   If the APA content will be reproduced in a book, then permission is limited to the life of the current edition.

6. Online Use Limitations.
   APA full-text articles or book chapters may only be posted on a secure and restricted web site. The credit line specified in Section 2 shall appear on the first screen on which the APA content appears and shall be amended to include the following additional wording: No further reproduction or distribution is permitted without written permission from the American Psychological Association.

7. Reservation of Rights.
   APA reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent on Payment.
   While you may exercise the rights licensed immediately upon issuance of the license at

the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by APA or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and APA reserves the right to take any and all action to protect its copyright in the materials. "$0.00" fee means that APA waived the fee.

9. License Revocation.
   APA or CCC may, within 60 days from the date of license, deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will APA or CCC be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to CCC for denied permissions.

10. Warranties.
    APA warrants that it has, to the best of its knowledge, the rights to license reuse of this content. However, you must ensure that the material you are requesting is original to APA and does not carry the copyright of another entity (as credited in the published version). If any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that original source to reuse the material.

11. Indemnity.
    You hereby indemnify and agree to hold harmless APA and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

12. No Transfer of License.
    This license is personal to you, but may be assigned or transferred by you to your employer or a business associate if you give prompt written notice of the assignment or transfer to the APA. No such assignment or transfer shall relieve you of the obligation to pay the designated license fee on a timely basis (although payment by the identified assignee can fulfill your obligation).

13. No Amendment except in Writing.
    This license may not be amended except in a writing signed by both parties (or, in the case of APA, by CCC on APA's behalf).

14. Objection to Contrary Terms.
    APA hereby objects to any terms contained in any purchase order, acknowledgment,

check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and APA (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

15. Jurisdiction.
    This license transaction shall be governed by and construed in accordance with the laws of the District of Columbia. You hereby agree to submit to the jurisdiction of the federal courts located in the District of Columbia for purposes of resolving any disputes that may arise in connection with this licensing transaction.

Other Conditions:

v1.2

Questions? customercare@copyright.com.

Permission request from Wolters Kluwer to reproduce John Loeser's onion skin model of pain

WOLTERS KLUWER HEALTH, INC. LICENSE
TERMS AND CONDITIONS
Apr 03, 2023

This Agreement between Caitlin Howlett ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	5521141482356
License date	Apr 03, 2023
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	Clinical Journal of Pain, The
Licensed Content Title	Pain and Suffering
Licensed Content Author	John Loeser
Licensed Content Date	Jun 1, 2001
Licensed Content Volume	16
Licensed Content Issue	2
Type of Use	Dissertation/Thesis
Requestor type	University/College
Sponsorship	No Sponsorship
Format	Print and electronic
Will this be posted online?	Yes, on an unrestricted website
Portion	Figures/tables/illustrations
Number of figures/tables/illustrations	1
Author of this Wolters Kluwer article	No
Will you be translating?	No
Intend to modify/change the content	No
Title	Cognitive flexibility and persistent pain
Institution name	University of South Australia
Expected presentation date	Apr 2023
Order reference number	0412
Portions	Figure 1 on page S3
Requestor Location	Caitlin Howlett 108 North Terrace

	Adelaide, 5001 Australia Attn: Caitlin Howlett
Publisher Tax ID	13-2932696
Total	0.00 AUD
Terms and Conditions	

Wolters Kluwer Health Inc. Terms and Conditions

1. **Duration of License:** Permission is granted for a one time use only. Rights herein do not apply to future reproductions, editions, revisions, or other derivative works. This permission shall be effective as of the date of execution by the parties for the maximum period of 12 months and should be renewed after the term expires.
    i. When content is to be republished in a book or journal the validity of this agreement should be the life of the book edition or journal issue.
    ii. When content is licensed for use on a website, internet, intranet, or any publicly accessible site (not including a journal or book), you agree to remove the material from such site after 12 months, or request to renew your permission license
2. **Credit Line:** A credit line must be prominently placed and include: For book content: the author(s), title of book, edition, copyright holder, year of publication; For journal content: the author(s), titles of article, title of journal, volume number, issue number, inclusive pages and website URL to the journal page; If a journal is published by a learned society the credit line must include the details of that society.
3. **Warranties:** The requestor warrants that the material shall not be used in any manner which may be considered derogatory to the title, content, authors of the material, or to Wolters Kluwer Health, Inc.
4. **Indemnity:** You hereby indemnify and hold harmless Wolters Kluwer Health, Inc. and its respective officers, directors, employees and agents, from and against any and all claims, costs, proceeding or demands arising out of your unauthorized use of the Licensed Material
5. **Geographical Scope:** Permission granted is non-exclusive and is valid throughout the world in the English language and the languages specified in the license.
6. **Copy of Content:** Wolters Kluwer Health, Inc. cannot supply the requestor with the original artwork, high-resolution images, electronic files or a clean copy of content.
7. **Validity:** Permission is valid if the borrowed material is original to a Wolters Kluwer Health, Inc. imprint (J.B Lippincott, Lippincott-Raven Publishers, Williams & Wilkins, Lea & Febiger, Harwal, Rapid Science, Little Brown & Company, Harper & Row Medical, American Journal of Nursing Co, and Urban & Schwarzenberg - English Language, Raven Press, Paul Hoeber, Springhouse, Ovid), and the Anatomical Chart Company
8. **Third Party Material:** This permission does not apply to content that is credited to publications other than Wolters Kluwer Health, Inc. or its Societies. For images

credited to non-Wolters Kluwer Health, Inc. books or journals, you must obtain permission from the source referenced in the figure or table legend or credit line before making any use of the image(s), table(s) or other content.

9. **Adaptations:** Adaptations are protected by copyright. For images that have been adapted, permission must be sought from the rightsholder of the original material and the rightsholder of the adapted material.
10. **Modifications:** Wolters Kluwer Health, Inc. material is not permitted to be modified or adapted without written approval from Wolters Kluwer Health, Inc. with the exception of text size or color. The adaptation should be credited as follows: Adapted with permission from Wolters Kluwer Health, Inc.: [the author(s), title of book, edition, copyright holder, year of publication] or [the author(s), titles of article, title of journal, volume number, issue number, inclusive pages and website URL to the journal page].
11. **Full Text Articles:** Republication of full articles in English is prohibited.
12. **Branding and Marketing:** No drug name, trade name, drug logo, or trade logo can be included on the same page as material borrowed from *Diseases of the Colon & Rectum, Plastic Reconstructive Surgery, Obstetrics & Gynecology (The Green Journal), Critical Care Medicine, Pediatric Critical Care Medicine, the American Heart Association publications and the American Academy of Neurology publications.*
13. **Open Access:** Unless you are publishing content under the same Creative Commons license, the following statement must be added when reprinting material in Open Access journals: "The Creative Commons license does not apply to this content. Use of the material in any format is prohibited without written permission from the publisher, Wolters Kluwer Health, Inc. Please contact permissions@lww.com for further information."
14. **Translations:** The following disclaimer must appear on all translated copies: Wolters Kluwer Health, Inc. and its Societies take no responsibility for the accuracy of the translation from the published English original and are not liable for any errors which may occur.
15. **Published Ahead of Print (PAP):** Articles in the PAP stage of publication can be cited using the online publication date and the unique DOI number.
    i. Disclaimer: Articles appearing in the PAP section have been peer-reviewed and accepted for publication in the relevant journal and posted online before print publication. Articles appearing as PAP may contain statements, opinions, and information that have errors in facts, figures, or interpretation. Any final changes in manuscripts will be made at the time of print publication and will be reflected in the final electronic version of the issue. Accordingly, Wolters Kluwer Health, Inc., the editors, authors and their respective employees are not responsible or liable for the use of any such inaccurate or misleading data, opinion or information contained in the articles in this section.
16. **Termination of Contract:** Wolters Kluwer Health, Inc. must be notified within 90 days of the original license date if you opt not to use the requested material.
17. **Waived Permission Fee:** Permission fees that have been waived are not subject to future waivers, including similar requests or renewing a license.
18. **Contingent on payment:** You may exercise these rights licensed immediately upon

issuance of the license, however until full payment is received either by the publisher or our authorized vendor, this license is not valid. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of Wolters Kluwer Health, Inc.'s other billing and payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

19. **STM Signatories Only:** Any permission granted for a particular edition will apply to subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and do not involve the separate exploitation of the permitted illustrations or excerpts. Please view: STM Permissions Guidelines
20. **Warranties and Obligations:** LICENSOR further represents and warrants that, to the best of its knowledge and belief, LICENSEE's contemplated use of the Content as represented to LICENSOR does not infringe any valid rights to any third party.
21. **Breach:** If LICENSEE fails to comply with any provisions of this agreement, LICENSOR may serve written notice of breach of LICENSEE and, unless such breach is fully cured within fifteen (15) days from the receipt of notice by LICENSEE, LICENSOR may thereupon, at its option, serve notice of cancellation on LICENSEE, whereupon this Agreement shall immediately terminate.
22. **Assignment:** License conveyed hereunder by the LICENSOR shall not be assigned or granted in any manner conveyed to any third party by the LICENSEE without the consent in writing to the LICENSOR.
23. **Governing Law:** The laws of The State of New York shall govern interpretation of this Agreement and all rights and liabilities arising hereunder.
24. **Unlawful:** If any provision of this Agreement shall be found unlawful or otherwise legally unenforceable, all other conditions and provisions of this Agreement shall remain in full force and effect.

For Copyright Clearance Center / RightsLink Only:

1. **Service Description for Content Services:** Subject to these terms of use, any terms set forth on the particular order, and payment of the applicable fee, you may make the following uses of the ordered materials:
    i. **Content Rental:** You may access and view a single electronic copy of the materials ordered for the time period designated at the time the order is placed. Access to the materials will be provided through a dedicated content viewer or other portal, and access will be discontinued upon expiration of the designated time period. An order for Content Rental does not include any rights to print, download, save, create additional copies, to distribute or to reuse in any way the full text or parts of the materials.
    ii. **Content Purchase:** You may access and download a single electronic copy

of the materials ordered. Copies will be provided by email or by such other means as publisher may make available from time to time. An order for Content Purchase does not include any rights to create additional copies or to distribute copies of the materials

Other Terms and Conditions:

v1.18

Questions? customercare@copyright.com.

Permission request from Taylor & Francis Group to reproduce Steve Berry's WCST scoring diagrams

Permissions Request

Type of use: Academic
Article title: Diagrammatic procedure for scoring the wisconsin card sorting test
Article DOI: 10.1080/13854049608406670
Author name: Steve Berry
Journal title: The Clinical Neuropsychologist
Volume number: 10
Issue number: 1
Year of publication: 1996
Name: Caitin Howlett
Street address: 108 North Terrace
Town: Adelaide
Postcode/ZIP code: 5001
Country: Australia
Email: caitlin.howlett@mymail.unisa.edu.au
Telephone: 0474776626
Intended use: To be used in your Thesis?
Are you requesting the full article?: No
Extract and number of words: Full diagrams including notes for each diagram; and permission to adapt the diagram
Details of figure/table: Diagram 1(a), Diagram 1(b), and Diagram 2
Which University?: University of South Australia
Title of your Thesis?: Cognitive flexibility and persistent pain
University repository URL:
Is this a "Closed" or "Open" deposit?: open
Additional comments:

Appendix F: Statements of Authorship

University of South Australia

Statement of Authorship

Title of Paper	Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility
Publication Status	Published
Publication Details Relevant reference details	*Miles, S., ***Howlett, C.A.,** Berryman, C., Nedeljkovic, M., #Moseley, G.L., & #Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, 1-9. https://doi.org/10.3758/s13428-021-01551-3 * Co-first authors # Co-last authors
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Caitlin A. Howlett
Contribution to the Paper Brief description of your work in thispublication	Review conception; literature analysis, interpretation and synthesis; drafting and revision of manuscript; approval of final manuscript; journal revisions.
Overall Percentage (%)	35%
Certification	This is an original publication containing research conducted for my PhD thesis.
Signature	[redacted] Date: 07/02/2023
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Stephanie Miles
% Contribution to the Paper	35%

Signature		Date: 07/02/2023
Name of Co-Author 2	Carolyn Berryman	
% Contribution to the Paper	7.5%	
Signature		Date: 07/02/2023
Name of Co-Author 3	Maja Nedeljkovic	
% Contribution to the Paper	7.5%	
Signature		Date: 09/02/2023
Name of Co-Author 4	G. Lorimer Moseley	
% Contribution to the Paper	7.5%	
Signature		Date: 06/02/2023
Name of Co-Author 5	Andrea Phillipou	
% Contribution to the Paper	7.5%	
Signature		Date: 07/02/2023

University of South Australia

Statement of Authorship

Title of Paper	Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults.
Publication Status	Published
Publication Details Relevant reference details	**Howlett, C.A.,** Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2021). Same room – different windows? A systematic review and meta-analysis of the relationship between self-report and neuropsychological tests of cognitive flexibility in healthy adults. *Clinical Psychology Review*, *88*(102061). https://doi.org/10.1016/j.cpr.2021.102061

PRINCIPAL AUTHOR

Name of Principal Author (Candidate)	Caitlin A. Howlett	
Contribution to the Paper Brief description of your work in thispublication	Study conception and design; literature screening; data collection, extraction, analysis and interpretation; drafting and revision of manuscript; approval of final manuscript; journal revisions.	
Overall Percentage (%)	50%	
Certification	This is an original publication containing research conducted for my PhD thesis.	
Signature	[redacted]	Date: 07/02/2023

CO-AUTHOR CONTRIBUTIONS

By signing the Statement of Authorship, each author certifies that:

I. the candidate's stated contribution to the publication is accurate (as detailed above);

II. permission is granted for the candidate in include the publication in the thesis; and

III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.

Name of Co-Author 1	Michael A. Wewege	
% Contribution to the Paper	5%	
Signature	[redacted]	Date: 06/02/2023

Name of Co-Author 2	Carolyn Berryman	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 3	Annika Oldach	
% Contribution to the Paper	3%	
Signature		Date: 06/02/2023
Name of Co-Author 4	Elizabeth Jennings	
% Contribution to the Paper	2%	
Signature		Date: 08/02/2023
Name of Co-Author 5	Emily Moore	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 6	Emma L. Karran	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 7	Kimberley Szeto	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 8	Leand	

% Contribution to the Paper	5%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 9	Stephanie Miles	
% Contribution to the Paper	5%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 10	G. Lorimer Moseley	
% Contribution to the Paper	10%	
Signature	[redacted]	Date: 06/02/2023

University of South Australia

Statement of Authorship

Title of Paper	Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis.
Publication Status	Published
Publication Details Relevant reference details	**Howlett, C.A.,** Wewege, M.A., Berryman, C., Oldach, A., Jenning, E., Moore, E., Karran, E.L., Szeto, K., Pronk, L., Miles, S., & Moseley, G.L. (2022). Back to the drawing board – The relationship between self-report and neuropsychological tests of cognitive flexibility in clinical cohorts: A systematic review and meta-analysis. *Neuropsychology*, *36*(5), 347-372. https://doi.org/10.1037/neu0000796

PRINCIPAL AUTHOR

Name of Principal Author (Candidate)	Caitlin A. Howlett	
Contribution to the Paper Brief description of your work in thispublication	Study conception and design; literature screening; data collection, extraction, analysis and interpretation; drafting and revision of manuscript; approval of final manuscript; journal revisions.	
Overall Percentage (%)	50%	
Certification	This is an original publication containing research conducted for my PhD thesis.	
Signature	[redacted]	Date: 07/02/2023

CO-AUTHOR CONTRIBUTIONS

By signing the Statement of Authorship, each author certifies that:

I. the candidate's stated contribution to the publication is accurate (as detailed above);

II. permission is granted for the candidate in include the publication in the thesis; and

III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.

Name of Co-Author 1	Michael A. Wewege	
% Contribution to the Paper	5%	
Signature	[redacted]	Date: 06/02/2023

Name of Co-Author 2	Carolyn Berryman	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 3	Annika Oldach	
% Contribution to the Paper	3%	
Signature		Date: 06/02/2023
Name of Co-Author 4	Elizabeth Jennings	
% Contribution to the Paper	2%	
Signature		Date: 08/02/2023
Name of Co-Author 5	Emily Moore	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 6	Emma L. Karran	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 7	Kimberley Szeto	
% Contribution to the Paper	5%	
Signature		Date: 07/02/2023
Name of Co-Author 8	Leander Pronk	

% Contribution to the Paper	5%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 9	Stephanie Miles	
% Contribution to the Paper	5%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 10	G. Lorimer Moseley	
% Contribution to the Paper	10%	
Signature	[redacted]	Date: 06/02/2023

University of South Australia

Statement of Authorship

Title of Paper	Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy
Publication Status	Published
Publication Details Relevant reference details	*__Howlett, C.A.,__ *Miles, S., Berryman, C., #Phillipou, A., & #Moseley, G.L. (2023). Conflation between self-report and neurocognitive assessments of cognitive flexibility: A critical review of the Jingle Fallacy. *Australian Journal of Psychology, 75*(1), 2174684. https://doi.org/10.1080/00049530.2023.2174684 * Co-first authors # Co-last authors
PRINCIPAL AUTHOR	
Name of Principal Author (Candidate)	Caitlin A. Howlett
Contribution to the Paper Brief description of your work in thispublication	Review conception; literature analysis, interpretation and synthesis; drafting and revision of manuscript; approval of final manuscript.
Overall Percentage (%)	35%
Certification	This is an original publication containing research conducted for my PhD thesis.
Signature	[redacted] Date: 07/02/2023
CO-AUTHOR CONTRIBUTIONS	
By signing the Statement of Authorship, each author certifies that: I. the candidate's stated contribution to the publication is accurate (as detailed above); II. permission is granted for the candidate in include the publication in the thesis; and III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.	
Name of Co-Author 1	Stephanie Miles
% Contribution to the Paper	35%
Signature	[redacted] Date: 07/02/2023
Name of Co-Author 2	Carolyn Berryman

% Contribution to the Paper	10%	
Signature		Date: 07/02/2023
Name of Co-Author 3	Andrea Phillipou	
% Contribution to the Paper	10%	
Signature		Date: 07/02/2023
Name of Co-Author 4	G. Lorimer Moseley	
% Contribution to the Paper	10%	
Signature		Date: 06/02/2023

University of South Australia

Statement of Authorship

Title of Paper	Investigating self-report and neuropsychological assessments of cognitive flexibility in people with and without persistent pain: An online, cross-sectional, observational study
Publication Status	Submitted for publication
Publication Details Relevant reference details	Under review at *The British Journal of Pain*

PRINCIPAL AUTHOR

Name of Principal Author (Candidate)	Caitlin A. Howlett	
Contribution to the Paper Brief description of your work in thispublication	Study conception and design; recruitment and data collection; analysis and interpretation; drafting and revision of manuscript; approval of final manuscript.	
Overall Percentage (%)	58%	
Certification	This is an original publication containing research conducted for my PhD thesis.	
Signature	[redacted]	Date: 07/02/2023

CO-AUTHOR CONTRIBUTIONS

By signing the Statement of Authorship, each author certifies that:

I. the candidate's stated contribution to the publication is accurate (as detailed above);

II. permission is granted for the candidate in include the publication in the thesis; and

III. the sum of all co-author contribution is equal to 100% less the candidate's stated contribution.

Name of Co-Author 1	Tyman Stanford	
% Contribution to the Paper	10%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 2	Carolyn Berryman	

% Contribution to the Paper	7%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 3	Emma L. Karran	
% Contribution to the Paper	5%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 4	Valeria Bellan	
% Contribution to the Paper	2%	
Signature	[redacted]	Date: 16/02/2023
Name of Co-Author 5	Scott Coussens	
% Contribution to the Paper	4%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 6	Stephanie Miles	
% Contribution to the Paper	4%	
Signature	[redacted]	Date: 07/02/2023
Name of Co-Author 7	G. Lorimer Moseley	
% Contribution to the Paper	10%	
Signature	[redacted]	Date: 06/02/2023