

Reverse Engineering language test
constructs for Messick's Value
Implications:
A Sociolinguistic Approach

Johanna Motteram

Bachelor of Arts, University of Adelaide, Australia

Master of Arts (Applied Linguistics), University of Adelaide, Australia

A thesis submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy

Discipline of Linguistics

University of Adelaide, Australia

March 2016

Table of Contents

Chapter 1	Introduction	14
1.1	Background	14
1.2	Aims and Questions	16
1.3	Theoretical and Methodological approaches	17
1.4	Thesis structure.....	18
1.5	Chapter Review	19
Chapter 2	Major Conceptual Framing of the Study.....	20
2.1	Language testing and language theory.....	21
2.2	High Stakes Language Testing.....	23
2.2.1	High Stakes English language testing in Australia.....	23
2.2.2	The Language Testing Industry	25
2.3	Assessing writing in high stakes tests	27
2.3.1	Writing a response to a task prompt	27
2.3.2	The social origins of the letter	28
2.3.3	The letter prompt and the text type ‘friendly letter’ in IELTS.....	33
2.3.4	Scoring criteria in writing in general	34
2.3.5	The relationship between scoring criteria and the test construct.....	34
2.3.6	Challenges of scoring criteria development and use.....	35
2.4	IELTS.....	36
2.4.1	IELTS General Training Version Written Module Task 1	37
2.4.2	Vocabulary used in task prompts.....	40
2.5	IELTS GTVWMT#1 Scoring Criteria.....	40
2.5.1	Task Achievement	41
2.5.3	Lexical Resource.....	46
2.5.4	Grammatical Range and Accuracy	47
2.5.5	Empirical and theoretical bases for the IELTS scoring criteria.....	49
2.6	Language theory, validity and the language testing literature.....	50

2.6.1	Perspectives and priorities of the two approaches to high stakes language testing ...	50
2.7	Introduction to the framework used in the study	53
2.8	Messick’s Validity Matrix and value implications; the hidden assumptions of the test construct	56
2.8.1	Messick’s validity matrix	56
2.8.2	Messick’s validity matrix and this study	59
2.8.3	Limitations of the approach to language which underpins the construct in most language assessment research	59
2.8.4	Performance tests, validity and scoring criteria	60
2.8.5	Messick as preferred foundation for language test validity discussion.....	62
2.9	A broad view of the location of the test	63
2.9.1	Security	63
2.9.2	History	63
2.9.3	Social, Geographical, and Political location of the test.....	66
2.9.4	Economic concerns	67
2.9.5	Education including washback.....	67
2.10	Responses to the socio political location of the test	68
2.10.1	Critical Language Testing literature	68
2.10.2	Calls for novel approaches.....	70
2.10.3	Why the value implications quadrant lacuna?	72
2.11	Bernstein’s pedagogic device and language test constructs	74
2.11.1	Bernstein and pedagogic discourse	75
2.11.2	Codes and the pedagogic device.....	76
2.11.3	The structure of the pedagogic device	80
2.11.4	Application of the pedagogic device in the study.....	84
2.12	Systemic Functional Linguistics, text analysis, and the realisation rules.....	86
2.12.1	Language is functional	86
2.12.2	Text is the basic unit of semantic analysis	87

2.12.3	Relationship of text and system.....	88
2.12.4	Relationship of text and context of situation	90
2.12.5	Metafunctions and register	91
2.13	Chapter Review	92
Chapter 3	Methodology and Method.....	94
3.1	Study Design Overview	94
3.2	Identification of target range of scripts	94
3.3	Ethics approval.....	95
3.4	Pre-data collection	95
3.4.1	Task selection.....	95
3.4.2	Survey development	96
3.5	Participant identification	96
3.5.1	Institutional engagement.....	96
3.5.2	Participant recruitment.....	97
3.6	Data collection sessions	97
3.6.1	Survey completion	97
3.6.2	Text composition.....	97
3.7	Rating procedures	98
3.7.1	Cambridge English Agreement.....	98
3.7.2	IDP Australia’s assistance with organising rating.....	99
3.7.3	Rating	99
3.8	Analytical Procedures	100
3.8.1	Under length scripts.....	100
3.8.2	Statistical Analyses.....	100
3.9	Qualitative Analyses of written scripts	102
3.9.1	Current conceptualisations of text complete-ness.....	103
3.9.2	Conceptualisation of a complete text.....	103
3.9.3	Analysis for determining text unity / Generic Structure Potential	105

3.9.4	Analyses of ‘tone’	106
3.9.5	Text features: ‘spoken like’	112
3.10	Chapter Review	114
Chapter 4	Analysis and results chapter	116
4.1	Chapter overview	116
4.2	Collected Texts and Survey Results.....	117
4.2.1	Data Set description	117
4.2.2	Survey results	117
4.2.3	Tabulated responses survey question 14	121
4.3	Rater Results and Statistical analyses	122
4.3.1	Rater Results	122
4.3.2	Preliminary organisation of the data	124
4.3.3	Cronbach’s Alpha	126
4.3.4	Multi-facet Rasch	127
4.3.5	Multi-facet Rasch and misfit texts	129
4.3.6	Bias interactions	129
4.4	Qualitative analyses of text: Field.....	131
4.4.1	Explain Lateness.....	131
4.4.2	Holiday locations.....	133
4.4.3	Sign offs and keep in touch messages	138
4.4.4	Naming.....	142
4.5	Qualitative analyses of text: Generic Structure Potential	144
4.5.1	Task mandated requirements of the task.....	144
4.5.2	Analytically identified additional Semantic text elements	145
4.5.3	Expanded investigation of semantic text elements.....	149
4.5.4	Order of text elements.....	157
4.5.5	Results of semantic text elements investigation	160
4.5.6	Adequate realisation of the semantic text elements	161

4.6	Qualitative analyses of text: Tenor	176
4.6.1	Whole text analysis for Tenor	176
4.6.2	Comparison tables for aspects of Tenor	180
4.7	Qualitative analyses of text: Mode	183
4.7.1	Reference to interactants	183
4.7.2	Lexical Density.....	184
4.7.3	Grammatical Intricacy / Projection	185
4.8	Chapter Review	186
Chapter 5	Discussion.....	187
5.1	Discussion of results of the statistical analyses	187
5.1.1	What do the Cronbach’s Alpha scores mean?	187
5.1.2	What do the bias interactions mean?.....	188
5.1.3	Does the TA score in Task 1 really matter anyway?	189
5.1.4	Should the TA score matter more?.....	190
5.2	Recognition Rules.....	192
5.2.1	Recognition Rules and the context of culture.....	192
5.2.2	Recognition rules and the contextual configuration of the task	194
5.3	Realisation Rules	197
5.3.1	Realisation Rules related to meeting the requirements / addressing the key bullet points of the task	198
5.3.2	Realisation Rules and the order of optional and obligatory semantic text elements	204
5.3.3	STEs, realisation rules, and the notion of an adequate realisation of the STE	206
5.3.4	Realisation Rules related to Tenor.....	214
5.3.5	Realisation rules and Field	218
5.3.6	Realisation rules and Mode	220
5.4	Alternative perspectives of the IELTS GTVWMT#1	221
5.4.1	The mystery of the missing task mandated STEs.....	222
5.4.2	An alternative perspective on the realisation rules, Politeness Theory	227

5.4.3	Confounding of the performance dimensions; restriction of clarity for analysts and potentially raters.....	228
5.4.4	Evaluative rules and test preparation.....	229
5.4.5	Candidate Voices in language testing research	231
5.4.6	Assumptions of the scoring criteria	233
5.5	Chapter Review	235
Chapter 6	Conclusion.....	237
6.1	A review of the research questions	237
6.2	General Findings	242
6.3	Contributions	244
6.4	Limitations.....	245
6.5	Recommendations	247
6.5.1	Recommendations for test developers.....	247
6.5.2	Recommendations for test takers and test preparation teachers	249
6.5.3	Recommendations for test score users	249
6.5.4	Recommendations for text book publishers.....	250
6.6	Recommendations for future research.....	250

Appendices

Appendix A	Immigration Document
Appendix B	Public Version of the Writing Band Descriptors for IELTS Task 1
Appendix C	Task prompt used in the study
Appendix D	Participant Survey used in the study
Appendix E	Proposal sent to Cambridge English for access to raters
Appendix F	Cambridge English / University of Adelaide agreement
Appendix G	Data set
Appendix H	Table of survey responses

- Appendix I Mood, Deixis, Polarity Table
- Appendix J Liz Bangs-Jones email to L-Test List
- Appendix K Top Tips for IELTS extract

References

Abstract

Industrial, high volume, high stakes language testing has recently emerged, which annually assesses the language proficiency of tens of millions of people, worldwide. Important decisions are made using test results, so the tests and the scoring procedures should be theoretically sound. This study explores an aspect of Messick's validity matrix which explicitly identifies the potential for validity threat where there are hidden social and cultural assumptions in the test construct. This facet of validity research and theorisation has been largely ignored by the language testing field.

The problem of implicit assumptions in test constructs is approached through a theoretical framework which combines Bernstein's model of the structuring of pedagogic discourse and the results of Systemic Functional Linguistics (SFL) informed text analysis. The framework allows connections between semantic and lexicogrammatical analysis of test responses and the social and cultural assumptions of the institution which evaluates that discourse. Analysis using SFL was pursued on the basis that the orientation to language which underpins the test construct under investigation, the International English Language Testing System (IELTS) General Training Version Written Module Task One (GTVWMT#1) is ill equipped to address questions of social and cultural assumptions implicit in texts, tasks, and test constructs.

The IELTS GTVWMT#1 scoring criteria performance dimension Task Achievement, and its constituent elements 'tone' and 'the requirements of the task' were the particular focus of the study. The embodiment relationship between a test's construct and the scoring criteria used to inform evaluation of test texts is an assumption of the study.

Responses (n=54) to a single IELTS GTVWMT#1 prompt were collected from a diverse potential candidate population. The prompt demanded a friendly letter which communicated three main messages. The responses were double marked by trained IELTS raters. Quantitative analyses informed targeted qualitative analyses, including text analysis to identify similarities between groups of texts with respect to semantic constituency, realisation of critical semantic text elements, and Tenor.

The study found that there are implicit expectations in the test construct related to social and cultural values and assumptions.

These come from;

1. The particular task used in the study. There is potentially limited access to the domains of language use specified in this task.
2. The situational context of testing which has an impact on need for explicitness, on order of information, and on inclusion of detail.
3. Community expectations of thank you letters, particularly as they are entwined with notions of gracious behaviour.

The major limitation of the study is the small sample size. Major recommendations include further work on the development of a model of text in social context to support language assessment, and consideration of the semantic constituency and potential for the impact of socially originating semantic variation in realisation of texts in high stakes language tests.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university of other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for another degree or diploma in any university of other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Johanna Motteram

Date

Acknowledgement of Help

Many thanks must go to my long suffering, patient, generous, and supportive Principal Supervisor, Dr Bill Winser. In addition, I acknowledge my other supervisors from the Discipline of Linguistics at the University of Adelaide who have assisted me over the years, Dr John Walsh, Dr Peter Mickan, and Dr Peter White.

I acknowledge the assistance and support of Cambridge English and IELTS Australia. Cambridge English allowing me access to trained raters, and for logistical assistance, Ms Jenny Osborne of IDP Australia. In particular, I acknowledge Dr Gad Lim of Cambridge English for ongoing support, timely responses, and for thoughtful feedback on a draft of this dissertation.

The community of scholars within the Discipline of Linguistics at the University of Adelaide, particularly Dr Jodie Martin, Jessica Scott, Dr Hiromi Teramoto, Dr Margareta Rebelos, Dr Nayia Cominos, and Dr Celine Chu.

I also acknowledge the help and forbearance of my family, my children Emma and Cooper who have grown up with me working on one thesis or another, and have developed skills in coffee supply and quiet avoidance, and my husband Phil who has supported me unstintingly. Finally my parents, Jill and Kim Read, who socialised me for life-long learning.

Preface

I once had a research meeting with an informant which changed my life. I was working on a project investigating candidate preparation for the International English Language Testing System (IELTS, an English language test) and my informant came to see me in a state of anger and confusion. She had been trying to get a high enough score in the IELTS General Training test to support her application for Permanent Residency in Australia. She needed a score of 7 overall, with no individual score below 7 (there are four scores reported: reading, writing, speaking and listening) but she was having trouble with the writing paper. She had attempted the test many times but could not score higher than 6.5 in the written section, even though she was a graduate of an Australian university and had been living and studying in Australia for more than five years. We talked about her writing and I asked to see a sample of it for the first task, a letter writing task. When I looked at her practice text alongside the scoring criteria I thought that she might be having problems with the scoring criteria section Task Achievement. Task Achievement includes the descriptors 'with the tone consistent and appropriate' (level 7), 'there may be inconsistencies in tone' (level 6), and 'the tone may be variable and at sometimes inappropriate' (level 5). The practice letter she showed me involved asking a landlord for assistance. Reading her text gave me the impression that she was angry and demanding. I was unsure but I felt that her response was probably not written with the level of courtesy and politeness that might be considered appropriate to the situation by whoever would be marking the test. We talked for a few minutes about how she could change her writing to make it sound more polite and she left.

She emailed me a few weeks later saying that her writing score had improved on her latest attempt at IELTS.

When I reported the incident to my project co-author, I recalled that throughout my association with the informant whenever we talked about writing she always focussed on the presence or lack of grammatical errors in her writing and then how her response answered the question. In time I realised that her understanding of test performance specifically and language use broadly may have been cast only in those terms.

Since that meeting I have thought quite a bit about what 'tone' might mean in the context of the IELTS Writing Task 1. I've also thought about scoring criteria for assessing writing and how language use and test performance is represented within them.