CHILDHOOD BEHAVIOUR PROBLEMS: DISCREPANCIES BETWEEN REPORTS FROM CHILDREN, PARENTS AND TEACHERS

VOLUME I

Michael Gifford Sawyer, MBBS (Monash), FRCPC, FRANZCP,
Dip Child Psych. (Toronto)

A thesis submitted in fulfilment of the
requirements of the degree of Doctor of Philosophy,
University of Adelaide.

Department of Psychiatry, University of Adelaide.

Submitted June, 1990.

For my mother, Dr Dilys Sawyer, and

my grandfather, Dr Robert Manners

INDEX

ABSTRACT

The aim of this thesis was to study the level of agreement between mothers, fathers, children, and teachers when describing childhood emotional and behavioural problems. In addition, possible reasons for the differences between reports from these informants were investigated. Finally, the pattern of differences between reports describing two different populations of children living in two-parent families were compared. The first population consisted of children living in the general community and the second of children who had been referred to mental health clinics.

In the general community, children consistently reported the most problems, teachers reported the fewest problems, and the number of problems reported by mothers and fathers was intermediate between the number reported by children and teachers. The number of children identified as "cases", on the basis of their exceeding the cutoff score on a child behaviour checklist, varied for different informants. The results of the study suggest that the degree of overlap between cases identified by different informants was little better than might be expected by chance.

Parent psychopathology and family adjustment showed a significant positive association with the number of childhood emotional and behavioural problems reported by parents. The size of the association varied for mothers and fathers, when describing male or female children. There were also significant differences in the pattern of discrepancies between parent and child reports describing children in the general

community, and reports describing children referred to mental health clinics.

The different number of problems reported by different informants makes it likely that the frequency with which children will be diagnosed as having emotional and behavioural disorders will vary, depending on the weighting given to reports from different informants. In addition, it appears that epidemiological studies which rely on reports from only one or two informants may fail to identify important groups of children with problems. Finally, the results of the study draw attention to the limitations of using only samples of children referred to mental health clinics when studying the pattern of discrepancies between parent and child reports of childhood emotional and behavioural disorders.

DECLARATION

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university and, to the best of my knowledge and belief, the thesis contains no material previously published or written by another person except when due reference is made in the text of the thesis. I consent to the thesis being made available for photocopying and loan if applicable if accepted for the award of the degree of Doctor of Philosophy.

Michael G. Sawyer

1st June, 1990

ACKNOWLEDGEMENTS

I wish to acknowledge the support of a number of people who helped with this thesis. Dr A. Pollard provided the initial encouragement and first laid down the challenge. Professor R. Kalucy helped to define the key questions addressed in the study and provided advice with regard to the methodology. This advice was supplemented by very helpful discussions with Professor A. Costello, Professor B. Nurcombe and Dr K. Baghurst. Professor M. Prior and Professor T. Achenbach both carefully read the thesis and provided a number of helpful suggestions. Judith White helped greatly with the typing of the thesis, and Johanna Gowland and Jenny Clark with the collection and coding of data. I am grateful to Professor R. Kosky for allowing me the opportunity to complete the thesis. Finally, I wish to express my gratitude to Dr P. Baghurst whose continuing supervision and encouragement was crucial for the success of the study.

PREFACE

Discrepancies between reports obtained from children, parents, and teachers are a critical problem hindering both research, and the accurate clinical assessment and treatment of childhood emotional and behavioural disorders. Typically, investigations of agreement between these informants have revealed low to moderate correlations in the order of $r < 0.3$ (Achenbach, McConaughty, & Howell, 1987). Currently there are no guidelines available to advise clinicians or researchers as to how they should weight discrepant reports in order to evaluate accurately childhood disorders. Thus, assessments made by clinicians and researchers will vary greatly, depending on the extent of agreement between different informants, and on whether information from one informant is weighted more heavily when disagreement exists. It is likely that the disagreement between informants contributes to unreliability in the diagnosis of childhood emotional disorders. This unreliability makes investigation of childhood disorders difficult, and may contribute to inconsistent clinical management of children.

Commenting on the problem of discrepant reporting, Herjanic and Reich (1982) have pointed out that it cannot be assumed that parental reports are more accurate than those obtained from their child simply because parents are older. Furthermore, both Hill (1985) and Herjanic and Reich (1982) have made the obvious point that children may be the only, or at least the most important, source of information about subjective distress involving feelings such as guilt, sadness, anxiety, or pain. Hill (1985) pointed out that, although personal distress may be inferred from public behaviour, parent reports may minimise the child's subjective suffering

because of ignorance or vested interest. In addition, emotional turmoil in the parents may affect their capacity to identify accurately their child's distress Herjanic and Reich (1982).

To date, there have been few studies which have attempted to identify factors which may be associated with the poor agreement between reports from different informants, and results to date are contradictory. For example, while Moretti, Fine, Haley, and Marriage (1985) reported that parent psychopathology may bias parent observations of their children, Weissman et al. (1987) have reported improved agreement when a parent informant was depressed at the time of the report. An important limitation of many studies which have investigated discrepancies in reports from different informants is that they included only children who were patients at mental health clinics (Jensen, Traylor, Xenakis, & Davis, 1988). As highlighted by Henderson, Byrne, and Duncan-Jones (1981), "Where one is looking for causation in the social or behavioural domains, it can be seriously misleading to study only clinical samples." (p. 85). This is because there are often substantial differences between the social and intrapersonal attributes of people with the same symptoms and disorders in clinic and community populations.

This thesis describes the results of a study which investigated discrepancies between reports obtained from children, parents, and teachers in two distinct samples of children. The first sample consisted of children randomly selected from the general community and the second consisted of children who had been referred to mental health clinics. The study focused on several possible factors which might contribute to discrepancies in the reports obtained from children and their parents in the general community. These factors included the age of the child, the

nature of the child's symptomatology, and emotional disturbance in the child's parent. In addition, the study compared the level of agreement between reports obtained from different informants describing children in the general community with the level of agreement between reports describing children referred to mental health clinics.

The unique features of the study are the combination of comprehensive data describing the participating children, obtained from four different informants in most instances, the inclusion of children in an age range where the reliability of their reports is acceptable (Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985), the inclusion of representative samples of children from the general community and of children who had been referred to mental health clinics, and finally, the use of the same measures to obtain reports about the children from the four different informants.

CHAPTER 1. INTRODUCTION

During the last decade there has been increasing interest in the efficacy of assessment methods used to evaluate children with emotional and behavioural problems, and in the validity and reliability of the schemas used to classify these problems.

Several factors appear responsible for the increased interest in assessment and taxonomy of childhood psychopathology. First, the President's Commission on Mental Health (1978) placed great emphasis on the need to gather reliable information on the incidence and prevalence of mental health problems, including those amongst children and adolescents. In order to achieve this, reliable and valid assessment methods were needed which could be used in community surveys.

Second, when the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) (American Psychiatric Association, 1980) was published, it described a large number of new disorders of childhood and adolescence. Unfortunately, initial excitement about the increased recognition given to disorders in young people was soon tempered by concern at the low reliability shown by clinicians when diagnosing the disorders (Rutter & Shaffer, 1980). In addition, doubts were expressed about the extent to which the disorders represented discrete entities with a unique aetiology, treatment response, and prognosis. For example, Bird et al. (1988) highlighted the limitations of the categorical approach employed in diagnostic schemas such as the DSM-III and ICD-9 (World Health Organization, 1978). Bird et al. (1988) suggested that

careful consideration needed to be given to methods which employ a dimensional rather than categorical approach when assessing childhood emotional and behavioural problems.

A third factor responsible for increased interest in the assessment and taxonomy of childhood problems was the development of new assessment methods which enabled the independent collection of information from children, parents and teachers. As highlighted by Kovacs and Beck (1977), "Description of a phenomenon and its reliable assessment are necessary first steps in order to build up a scientific body of knowledge" (p. 23). The development of new assessment methods facilitated the necessary first steps needed to achieve a more comprehensive description of childhood problems. These methods included two different, but complementary approaches. The first approach consisted of structured interviews (Herjanic, 1984) which could be administered independently to children and parents. The second consisted of checklists or questionnaires (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986; Achenbach & Edelbrock 1987) which could be completed independently by children, parents, and teachers.

The availability of new assessment methods encouraged studies which compared reports of children's problems obtained from different informants. These studies revealed only low to moderate agreement between informants and also highlighted the value of children as informants about their own emotional and behavioural problems (Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985). The results renewed interest in the general problem of how information obtained from different informants, often showing little agreement, could be integrated during the assessment of childhood emotional and behavioural disorders. As

pointed out by Edelbrock and Costello (1988), a problem "with the DSM is that it specifies *what* to assess but not *how* to assess it" (p. 221). At present, there are no generally accepted rules which can be employed in order to resolve disagreements between informants when attempting to establish a child's diagnosis or treatment plan.

For the purpose of this review, all relevant studies were initially stratified into categories which were representative of the different approaches used to investigate the level of agreement between informants. Subsequently, studies were selected for inclusion in the review on the basis that they best illustrated the type of research being discussed within each category.

Relevant studies were identified from a computer search of Psychological Abstracts from 1980 to 1989, and from the comprehensive review by Achenbach, McConaughy, and Howell (1987). In addition, manual searches were made of the following journals from 1980 to 1989: Journal of Abnormal Child Psychology, Journal of the American Academy of Child and Adolescent Psychiatry, Journal of Child Psychology and Psychiatry, American Journal of Orthopsychiatry, Child Development, Monographs of the Society for Research in Child Development, and the Journal of Consulting and Clinical Psychology.

1.0 BACKGROUND

A difficulty in comparing the results from studies which have investigated the level of agreement between informants describing

children's behaviour is that the studies vary in a number of key areas. These include:

a) The Study Sample: The samples of children participating in studies have varied. Different studies have included children attending mental health clinics (Edelbrock, Costello, Dulcan, Conover, & Kalas, 1986), children whose parents had a psychiatric illness (Weissman et al., 1980) children attending paediatric clinics (Herjanic & Reich, 1982), and children in the general community (Jensen, Traylor, Xenakis, & Davis, 1988). It is plausible that these differences may explain, in part, some of the differences between the results found in the studies as there are often substantial differences between the social and intrapersonal attributes of people with the same symptoms and disorders in clinic and community populations (Henderson, Byrne, & Duncan-Jones, 1981).

b) The Method of Classifying Children's Problems: Studies have varied in the approaches they have used to classify children's emotional and behavioural problems. Both "dimensional" and "categorical" approaches have been employed. The dimensional approach assumes that childhood problems lie on a continuum extending from normality to severe disturbance. Studies which employed this approach focused on the number of problems reported by different informants, and the level of consistency between informants (Achenbach et al., 1987). In these studies, information about children's behaviour was generally collected by means of self-report questionnaires (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986; Achenbach & Edelbrock, 1987). The categorical approach involves classifying children into one of two groups containing either children identified as having disorders, or those without disorders. In addition, children may be further categorized as

having different types of disorders. Examples of this type of approach include studies in which comparisons were based on psychiatric diagnoses (Reich, Herjanic, Welner, & Gandhy, 1982), and studies which focused on the level of agreement about whether or not children scored above established cutoff scores on self-report questionnaires (Graham, 1967; Hulbert, Gdowski, & Lachar, 1986). In these studies information was collected by means of separate interviews with different informants or by using self-report questionnaires.

c) The Statistical Analyses: The statistical analyses employed to compare reports obtained from different informants have varied. Jacob, Grounds, and Haley (1982) have drawn attention to the wide range of different statistics which have been used in studies investigating the level of agreement between different informants. As well, Jacob et al. (1982) highlighted the different focus of these different statistical techniques. For example, correlation coefficients have been widely used as a measure of agreement between informants reporting on children's behaviour (Achenbach et al., 1987). As such, they provide a measure of the strength of the linear association between two sets of scores. However, as pointed out by Jacob et al. (1982), correlation coefficients do not provide information about the level of agreement on specific items, nor do they provide information about the relative magnitude of the scores reported by different informants. These are important considerations when evaluating the level of agreement between reports obtained from children, parents, and teachers. Finally, it should be noted that the number of subjects included in different studies has varied greatly. This is important because the number of subjects in a study has a major influence on its statistical power.

d) The Childhood Problems Studied: The types of childhood problems investigated in different studies have varied. For example, increased interest in childhood depression was an important influence which resulted in a number of studies focusing on this disorder (Moretti, Fine, Haley, & Marriage, 1985; Kazdin, French, & Unis, 1983; Kazdin, French, Unis, & Esveldt-Dawson, 1983). More recently, a series of reports by Fergusson and Horwood (Fergusson & Horwood, 1987a, 1987b, 1989) focused specifically on childhood conduct disorders. In light of this, it is unclear whether the results reported for the children with conduct disorders are applicable for children with other disorders.

e) The Age of Participating Children: Studies have varied greatly in the age of participating children. For example, the studies by Weissman and colleagues (Weissman et al., 1987; Angold et al., 1987) included a large number of "children" who were aged 19 years or more. In contrast, other studies focused on much younger children (Herjanic & Reich, 1982; Reich et al., 1982). This is an important issue because Edelbrock et al. (1985) have reported that the reliability of interviews with children under 10 years of age is relatively low. As a result, they suggested that the results of structured interviews with children under the age of 10 years should be interpreted with caution.

f) The Measures: The measures employed to obtain reports from different informants have varied in different studies. For example, some studies have used structured interviews (Herjanic & Reich, 1982; Reich et al., 1982) while others have used semi-structured interviews (Verhulst, Althaus, & Berden, 1987) or self-report measures (Kazdin, French, & Unis, 1983, Kazdin, French, Unis, Esveldt-Dawson, 1983; Kazdin, Esveldt-Dawson, Unis, & Rancurello, 1983). As well, within individual studies, there are

differences in the extent to which informants used the same type of measure when reporting on children's behaviour. Some studies have employed quite different measures with different informants (Anderson, Williams, McGee, & Silva, 1987) and this makes it difficult to know whether differences between reports are due to differences in the measures used, or differences in the characteristics of the informants.

g) The Informants: The number of informants included in different studies have varied. However, most reports described results for only a single pair of informants, for example, mothers and children (Herjanic & Reich, 1982; Edelbrock et al., 1986), or mothers and fathers (Ferguson, Partyka, & Lester, 1974; Jacob et al., 1982).

Finally, it is important to note that some studies have included all the children in each participating family, thereby having a single parent in the study report on several children. Other studies included only one child and one parent from each participating family.

The first part of this chapter reviews studies which attempted to identify factors which may influence reports obtained from individual informants. The second part reviews studies which investigated the level of agreement between pairs of informants. The final section highlights the specific problems created by discrepancies between informants for epidemiological studies which investigate the prevalence of childhood disorders.

1.1 STUDIES INVESTIGATING FACTORS WHICH INFLUENCE REPORTS FROM INDIVIDUAL INFORMANTS

This section reviews studies which identified factors which may influence reports obtained from individual informants.

1.11 Factors influencing reports from Children

Early comments describing factors which might influence children's reports were largely anecdotal and generally made in the context of discussing the development of new self-report questionnaires. For example, Wirt and Broen (1956) suggested that "willingness to say deviant things about the self" (p. 482) might influence children's reports about their own anxieties. Cowen, Zax, Klein, Izzo, and Trost (1965), on the other hand, suggested that "willingness of the respondent to admit to certain symptomatology or to ascribe socially undesirable characteristics to himself" (p. 685) may be less of a problem with children who "may be more inclined to accept questions at face value and who may be less aware of cultural stereotypes about certain types of undesirable self-referent statements" (p. 686).

The study of factors which may influence children's reports was hindered by a lack of widely accepted and reliable assessment methods suitable for use with children. For example, when describing the reliability of the semi-structured interview used in the Isle of Wight study, Rutter and Graham (1968) noted that "Surprisingly little has been written on the interview with the child as a diagnostic tool, in contrast to the

voluminous literature on the therapeutic interview." (p. 563). Further, they suggested that where interviews for children had been developed for the purpose of diagnosis, they tended to have focused on the assessment of unconscious conflicts and fantasies. Similar comments were made by Herjanic, Herjanic, Brown, and Wheatt (1975), who noted that interviews with children had tended to be "accepted with a generous portion of scepticism" (p. 41) and numerous roundabout techniques such as drawings, puppets, and games had been devised to facilitate the collection of information from children. Herjanic et al. (1975) pointed out that an unfortunate consequence of this was that it led to the development of child interviews which had "little value in terms of data collection and research" (p. 41).

In the late 1960s and early 1970s a number of interviews suitable for use with children began to be developed. For example, Rutter and Graham (1968) developed the semi-structured interview which was used to evaluate children in the Isle of Wight study. They investigated various aspects of the validity and reliability of the interview and concluded that it was "a reasonably sensitive diagnostic instrument which could give rise to reliable and valid judgments on whether the child exhibited any psychiatric disorder" (p. 576). However, they cautioned that reliability of the interview was more mediocre for specific symptoms, especially those covering the areas of anxiety and depression.

Herjanic et al. (1975) described the early development of a structured interview suitable for use with children and parents. They used the interview to compare reports from 50 children, aged 6-16 years who were attending a mental health clinic, and their parents. Herjanic et al. (1975) reported that the "ratio of agreement between parents and children

compares favourably to those between adults under similar circumstances .... We conclude, therefore, that children are reliable reporters and the use of a structured interview with children is worthy of further study." (p. 47).

It has been suggested that various factors may influence reports from children. These have included social desirability (Seibert & Ramanaiah, 1978), age and cognitive development (Kazdin, & Petti, 1982), and a failure to acknowledge "behavior which has repeatedly drawn reprimands from adults" (Poznanski, Cook, & Carroll, 1979). However, there have been few studies which have attempted to determine systematically the influence of these factors on children's reports.

One study which did investigate factors which may effect the reliability of reports from children was that by Edelbrock et al. (1985). This was one of a group of studies (Costello, Edelbrock, Dulcan, Kalas, & Klaric, 1984; Edelbrock et al., 1985; Edelbrock et al., 1986, Edelbrock & Costello, 1988) which described the development of the DISC. The DISC is a structured interview which was developed under contract for the National Institute of Mental Health for use in epidemiological surveys. Parallel interviews are available for use with children (DISC) and their parents (DISC-P). The study (Edelbrock et al., 1985) included 242 children, aged 6-18 years, who had been referred to a mental health clinic and focused on whether children's age, and the type of symptoms being described, influenced the test-retest reliability of the Diagnostic Interview for Children and Adolescents (DISC). Results from the study suggested that reliability was higher for "behavior and conduct symptoms than affective and neurotic symptoms" (p. 269) and "reliability of the child's report increased with age in almost all symptom areas" (p. 269).

Interestingly, Edelbrock et al. (1985) found that the children reported "substantially fewer and less serious symptoms during the second interview" (p. 271). This occurred for children of all ages but was most pronounced for the younger children. The phenomenon has been documented with other informants and has been attributed to practice effects and statistical regression. Edelbrock et al. (1985) reported that their interviewers felt that some children may have reported fewer symptoms in order to reduce the length of the second interview.

In concluding, Edelbrock et al. (1985) cautioned that the reliability of the DISC interview with children was substantially lower than that of the parallel interview designed for use with parents (DISC-P). This was particularly evident with the younger children. In light of this, they recommended that the results of structured interviews with children under 10 years of age be interpreted with caution. This is an important issue because many of the studies investigating the level of agreement between parents and children have included children who were less than 10 years old. As such, it is possible that the results of these studies were influenced, in part, by the lower reliability of the children's reports.

1.12 Factors influencing reports from Parents

Two related approaches have been used to try and identify factors which may bias parent reports and both have been hindered by the lack of a "gold standard" against which the accuracy of parent reports of childhood behaviour can be compared (Bird et al., 1987). The first approach used reports from an independent rater who observed the child's behaviour,

usually in the child's home. These observations were then used as a "gold standard" against which parent reports could be compared.

The second approach used referral to a mental health clinic as a criterion against which parent reports were compared. The use of this approach developed from the observation that parents play a critical role in the process of a child's referral to a mental health clinic. As highlighted by Ferguson et al. (1974), "A child is usually initially identified as deviant, maladjusted , or 'having problems' by the significant others in his life .... it is generally parental concern which provides the motivation for evaluation and treatment in child-guidance settings. Thus it is important to assess parental perceptions in order to understand the referral and diagnostic process." (p. 169).

Forehand and his colleagues have made extensive use of direct observation in order to investigate factors which may affect parents perceptions of children's behaviour (Forehand, King, Peed, & Yoder, 1975; Griest, Wells, & Forehand, 1979; Griest, Forehand, Wells, & McMahon, 1980; Rickard, Forehand, Wells, Griest, & McMahon, 1981; Forehand, Wells, McMahon, Griest, & Rogers, 1982; Forehand & Furey, 1985; Middlebrook & Forehand, 1985; Forehand, Lautenschlager, Faust, & Graziano, 1986; Furey & Forehand, 1986; Furey, Forehand, Baskin, & Tauber, 1986; Brody & Forehand, 1986; Forehand, Brody, & Smith, 1986; Long, Forehand, Fauber, & Brody, 1987). Initially, the studies focused on differences in the perceptions of mothers of clinic-referred children and non-clinic children (Forehand et al., 1975), subsequently the focus of the studies shifted to an examination of factors, such as maternal depression and marital disharmony, which may influence parents' perceptions of their child's behaviour (Griest et al., 1979; Rickard et al., 1981; Forehand et

al., 1982; Middlebrook & Forehand, 1985; Furey & Forehand, 1986; Furey et al., 1986; Brody & Forehand, 1986; Forehand et al., 1986).

A key element in the program of studies by Forehand and his colleagues was the use of direct observation by an independent rater as a method to assess children's behaviour. This involved training observers who visited the child's home on several occasions and recorded specific behaviours such as compliance to a request, noncompliance, and other deviant behaviour (whining, crying, tantrums, aggression). This methodology has attracted considerable attention and several authors have suggested that observations recorded during home visits need to be interpreted cautiously. For example, Lobitz and Johnson (1975) suggested that parents may bias home observations by "manipulating the target child to appear socially desirable or undesirable .... parents of referred children could have been manipulating their child's behavior in a socially undesirable way to guarantee treatment, whereas parents of nonreferred children might have been manipulating their child's behavior in a socially desirable direction to validate their report of 'normality'." (p. 370). Schaughency and Lahey (1985) pointed out that lack of consistency between direct observation and maternal reports may be due to the different approaches used when rating the children's behaviour, "i.e. direct observations of discrete behaviors during relatively brief intervals of time as opposed to more global ratings of child behavior that reflect broader knowledge of child behavior" (p. 718). Achenbach et al. (1987) have drawn attention to the fact that direct recording of behaviour by trained observers is "further limited by practical considerations such as cost, as well as by the reactivity, low base rate, and covert nature of many important problems" (p. 214). Finally, it is important to note that the studies by Forehand et al.

(1975; 1982; 1985; Forehand, Lautenschlager, Faust, & Graziano, 1986; Forehand, Brody, & Smith, 1986) focused on children aged 2-9 years, the behaviour studied was largely noncompliance, and the parent was nearly always a mother.

The key finding reported in the studies by Forehand and his colleagues was that maternal perception of children's behaviour was influenced by factors other than the behaviour of the child (as rated by home observation). For example, in a study of 22 clinic-referred children, aged 3-8 years, Griest et al. (1979) reported that maternal depression was a significant predictor of mothers' perceptions of childhood maladjustment. Subsequent studies (Middlebrook & Forehand, 1985; Forehand, Brody, & Smith, 1986) reported that stress and marital dissatisfaction were also associated with maternal ratings of child deviance. Finally, in their most recent papers, the group argued that "parental perceptions are caused by no single factor; rather, our approach can be summarized in terms of the combined influences of parent characteristics. This analysis of the determinants of parental perceptions of maladjustment suggested that particular characteristics of parents, such as depression, combine with behavioral characteristics of the child to produce perceptions of maladjustment" (Brody & Forehand, 1986, p. 238). As mentioned, in these studies maladjustment was defined largely as non-compliant behaviour, however, a recent study by Conrad and Hammen (1989) has reported similar results in a study of maternal perceptions of childhood depressive symptoms and other internalizing symptoms.

A good example of the second approach used to study factors which may influence parent reports is the epidemiological study of Buckinghamshire

children in England (Mitchell & Shepherd, 1966; Shepherd, Oppenheim, & Mitchell, 1966). In the study, Mitchell and Shepherd (1966) sent questionnaires "dealing with behaviour, health and family background" (p. 249) to parents of a one-in-ten random sample of children, aged 5-15 years, attending school in Buckinghamshire. Questionnaires covering "attendance, attainment and physical disabilities as well as the presence of behaviour 'problems' in school" (p. 249) were sent to the child's teacher. The total sample consisted of 6300 children and very high response rates were obtained from parents and teachers.

In their study, Shepherd et al. (1966) noted that many "supposedly normal" children living in the community had problems which appeared identical to those commonly found amongst children attending mental health clinics. This was of concern because it was not clear why some children with problems were being referred to clinics and others, with similar problems, were not being referred for help. In order to investigate the phenomenon, a comparison was made between 50 children attending a mental health clinic and 50 children exhibiting similar behaviour who had never attended a mental health clinic. Shepherd et al. (1966) focused on several plausible factors which might have differentiated the groups. These factors included geographical distance from the clinic, presence of young children in the household, employment status of the mother, and " ... parental reactions to the children's behaviour. Of these factors, only parental reactions distinguished between the two groups" (p. 43).

From the results of their study, Shepherd et al. (1966) concluded "The mothers of clinic-children were more apt to be anxious, depressed and easily upset by stress; they were less able to cope with their children,

more apt to discuss their problems and to seek advice." (p. 47). This finding suggested that the characteristics of the mothers were playing a role in the way they perceived their children's behaviour and in their decision to seek referral to a mental health service. This is an important study because it reached a similar conclusion to that of Forehand, Lautenschlager, Faust, & Graziano (1986) while using a different methodology. More recently, Fergusson and Horwood (1987) have made similar observations, stating "It seems likely that independently of the child's behavioural tendencies, mothers who are depressed or under stress may see their children as being more troublesome and prone to conduct disorder (or that children with mothers who are depressed or stressed tend to develop situation specific conduct disorders)" (p. 269).

Several recent studies (Moretti et al., 1985; Schaughency & Lahey, 1985; Weissman, et al., 1987; Jensen, Traylor, Xenakis, & Davis, 1988; Jensen, Xenakis, Davis, Degroot, 1988; Webster-Stratton, & Hammond, 1988; Conrad & Hammen, 1989) have suggested that parental psychopathology may influence parents' perceptions of their children's behaviour. For example, Moretti et al. (1985) studied the level of agreement between 60 children, aged 8-17 years, and their parents when reporting on the presence of depressive symptoms in the child. Children in the study had been referred for assessment of affective symptoms and were either inpatients or outpatients at a child psychiatry clinic. Children and parents completed the Children's Depression Scale (CDS & CDS-P) (Tisher, & Lang, 1983) and parents also reported on their own depressive symptoms using the Beck Depression Inventory (Beck, Rush, Shaw, & Emery, 1979). A key finding in the study was that "The relationship between parent-ratings of depression in themselves on the BDI and their perceptions of depression in their children (CDS-P total depression) was significant

($r(32)=0.44, p<0.02$). However, parent-ratings of their own depression did not correlate significantly with their children's self-reports of depression on the CDS (total depression) ($r(32)=0.25$, NS) or the CDI ($r(32)=-0.08$, NS)" (p. 301). Moretti et al. (1985) were cautious when interpreting the meaning of this result but did suggest that " ... Perhaps depressed parents project their feelings of dysphoria onto their children and thus are biased in their evaluations" (p. 302).

A study by Jensen, Traylor, Xenakis, and Davis (1988) focused on the level of agreement between a group of 108 children, aged 6-11 years, and their parents. All the fathers in the families were officers or senior enlisted personnel with the U.S. Army. In the study, different measures were completed by the children and the parents. The Child Behavior Checklist (CBCL) (Achenbach & Edelbrock, 1983) was used to collect information from the parents while the children completed the Child Depression Inventory (Kovacs, & Beck, 1977) and the Children's Manifest Anxiety Scale (Reynolds, & Richmond, 1978). A modified version of the Hopkins Symptom Checklist (Derogatis, Lipman, Rischels, Uhlenhuth, & Covi, 1974) was used to measure the psychological adjustment of the parents. Consistent with the findings reported by Moretti et al. (1985), the parent scores (both mother and father) on the HSCL showed significant correlations with the total problem score, externalizing score, and internalizing score on the CBCL completed by the parents. These results in a community population suggested that the results reported by Moretti et al. (1985) were not unique to children referred to clinics for assessment of affective symptoms. In addition, the results from the study by Jensen, Traylor, Xenakis, and Davis (1988) suggested that parent psychopathology was significantly associated with parent reports of both externalizing and internalizing symptoms in their children. However,

when discussing their findings, Jensen, Traylor, Xenakis, and Davis (1988) made the important point that although the findings " ... may suggest a distortion or bias in parents' reports of their children based on their own symptoms, an equally plausible explanation is that parents' and children's symptoms are in fact related (through either genetic or environmental factors)." (p. 446). In other words, it is possible that there may be no bias in the parent reports, instead they accurately reflect an association between the presence of childhood psychological problems and parent psychopathology.

A final intriguing suggestion by Jensen, Traylor, Xenakis, and Davis (1988) is that parents' sex may have an influence on their ratings of childhood behaviour and that interparent reliability may " ... be related to the gender of the rating parent and the rated child. The fact that parents show fairly good reliabilities when rating sons (despite significant differences in the actual magnitude of their CBCL ratings) may indicate a higher threshold in fathers than mothers in perceiving sons' behavioral problems. Alternatively, mothers may be unduly sensitive and overreport sons' problems" (p. 445). Schaughency and Lahey (1985) have made a similar suggestion noting that "factors involved in paternal perceptions of child deviance may be quite different from maternal perception" (p. 721).

Finally, as well as reporting on the reliability of DISC interviews with children, Edelbrock et al. (1985) also reported on the test-retest reliability of the DISC-P employed with parents. Although the reliability of the reports from parents describing children of different ages varied less than that observed with the reports from children, " ... higher reliabilities were generally found for parents of younger children

than for parents of older children" (p. 270). Edelbrock et al. (1985) suggested that this pattern may have been due to parents having less opportunity to observe the behaviour of the older children because the children spend more time away from the family home. There was also a consistent decline (statistically significant in the case of the 10-13 year old children) in the number of parent-reported symptoms for children of all ages. This was a similar pattern to that found with the child reports, however, the magnitude of the decline of parent scores was smaller than that found with the children.

1.13 Factors influencing reports from Teachers

There have been relatively few studies looking at factors which may influence teacher ratings of children's behaviour. Touliatos and Lindholm (1975) studied teacher reports of 1,999 children, from kindergarten to grade 5, in regular classes in suburban schools in Texas. The Behavior Problem Checklist (BPCL) (Quay, & Peterson, 1967) was used to collect teacher ratings. Interestingly, the investigators reported that "differences between schools and between teachers were responsible for more of the variance on the BPCL than grade, sex, and social class." (p. 124). It is possible that this was due to marked differences in the prevalence of problems in different classes and schools. Alternatively, there may have been differences in the way participating teachers perceived childhood behaviour problems. Unfortunately, no reports were available from other informants with which the teachers' reports could be compared. As a result, it is unclear which is the more likely explanation of the findings from the study.

The specific characteristics of different approaches which can be used to assess classroom behaviour have been explored by Kazdin, Esveldt-Dawson, and Loar (1983). In their study, Kazdin et al. (1983) compared teacher reports with reports obtained from a "rater" who completed the same measures as the teachers after observing children's classroom behaviour, and with an "observer" who scored the presence or absence of on-task, disruptive, or neutral behaviour during specified time intervals. It was hypothesized that there would be better rater-observer agreement than teacher-observer agreement because the raters and observers would be able to observe the children more closely than the teacher. The study involved 32 children, aged 7-13 years, who were inpatients in a psychiatric intensive care facility. The teachers and the raters independently completed three measures of classroom behaviour: the Conners Teacher Rating Scale (Conners, 1969), the Teacher Report Form (Achenbach & Edelbrock, 1986) and an "Estimate of Overt Classroom Behavior" (p. 553). As predicted, rater-observer agreement was consistently higher than teacher-observer agreement. As well, it was noted that the reports from the three informants correlating with each other only "in the low to moderate range" (p. 561).

Kazdin et al. (1983) suggested that the limited agreement between the different informants in their study may have reflected differences in the focus of the assessment methods employed by the observers and the other informants, and differences in the opportunity teachers have to observe specific children in their classroom. They suggested that "Measures from different sources may vary in their utility for purposes of initial assessment, diagnosis, or delineation of groups and for evaluation of changes in specific behaviors over time." (p. 563). Teacher reports appeared useful in delineating particular diagnostic groups, such as

hyperactivity, while direct observation may be more useful for tracking specific behaviours over time. Costello and Angold (1988) have also highlighted the evidence which suggests that teachers focus on conduct or aggression problems and are less sensitive to internalizing problems

1.2 STUDIES INVESTIGATING THE LEVEL OF AGREEMENT BETWEEN INDIVIDUAL INFORMANTS

This section reviews studies which have reported the level of agreement between pairs of informants when describing emotional and behavioural problems in children.

For each pair of informants, individual studies are grouped together on the basis of the sample of children included in the study (children referred to mental health clinics, children with psychiatrically disordered parents, and children living in the general community), and the method employed to obtain reports from different informants (interview or self-report measure).

1.21 STUDIES OF AGREEMENT BETWEEN PARENTS AND CHILDREN

This section reviews studies which have reported the level of agreement between parents and children when describing emotional and behavioural problems in the children.

1.211 Early studies

In the early 1960s, the development of new self-report measures for children led to the recognition that there was only moderate agreement between reports completed by children and their parents. For example, Hafner, Quast, Speer, and Grams (1964) compared reports describing childhood anxiety obtained from children and their parents. The modified Children's Manifest Anxiety Scale and the General Anxiety Scale for Children were administered to the children. Modified versions of the same questionnaires were administered to the children's parents. The study included two groups of forty children, aged 9-12 years. One group had been referred to a psychiatric clinic and the other to a paediatric clinic. The authors reported that there was only "a moderate amount of agreement shown between the children's own self-ratings of anxiety and those of their parents" (Hafner et al., 1964, p. 557). In addition they reported that the children consistently identified more problems than their parents. Hafner et al. (1964) suggested that these findings reflected a lack of awareness by parents of their children's feelings and they recommended that the discrepancies between reports obtained from children and their parents needed considerably more investigation. Interestingly however, although later studies continued to note that discrepancies existed between reports obtained from children and parents, the issue received only limited attention until the last decade.

1.212 Studies of Children referred to Mental Health Clinics

1.2121 Studies using Structured Interviews

In 1982, Herjanic and Reich (Herjanic & Reich, 1982; Reich et al, 1982) described the level of agreement between 307 children, aged 6-16 years, and their mothers. Information about the children was obtained by means of a structured diagnostic interview which was administered independently to both informants. All the participating children had been referred to either a psychiatric clinic, a psychiatric consultation service, or a paediatric clinic. The level of agreement between the children and mothers was compared for both individual symptoms in a range of areas, and for the diagnoses derived from the separate interviews. The kappa statistic (Cohen, 1960) was used to describe the level of agreement between the different informants and the McNemar chi-square test (Armitage & Berry, 1987) was used to identify "asymmetrical reporting". The latter was defined as a situation where one informant reported significantly more symptoms than the other. The investigators pointed out that "percentage agreement" is an unsatisfactory index of agreement between informants describing rare symptoms because high levels of agreement are likely by chance alone (when both informants agree that a symptom is absent).

Herjanic and Reich (1982) described four characteristics which were associated with symptoms or behaviours which showed good agreement between the children and mothers:

1. The symptoms were identified by "objective, concrete questions" (p. 311).
2. The symptoms concerned "behavior reflecting a degree of seriousness that could not be overlooked" (p. 311).
3. The symptoms concerned "problems or events which would not likely be misunderstood or misinterpreted by children" (p. 317).
4. The symptoms were by their nature "likely to be known to the mothers" (p. 317).

Two characteristics were associated with symptoms showing poor agreement:

1. The symptoms "required judgement as to the presence or absence of a symptom" (p. 319).
2. The symptoms were identified by questions which "could easily be misunderstood or misinterpreted" (p. 319).

Four characteristics were associated with symptoms which were reported more frequently by the children than their mothers:

1. The symptoms were "of a very subjective nature, including those of various neurotic disorders and depression" (p. 320).
2. The symptoms described "somatic symptomatology" (p. 320).
3. The symptoms described "antisocial behavior, sometimes quite severe in nature" (p. 320).
4. The symptoms described "psychotic symptoms" (p. 321).

Finally, symptoms which were reported more frequently by mothers were characterized as "behavior that might be troublesome to the mother, or

described personality characteristics of the children that perhaps were more obvious to the mothers than the children themselves" (p. 321). In concluding, Herjanic and Reich (1982) suggested that further research was needed to investigate the influence of personality disturbance or psychiatric disorder in the parent, on parent reports of childhood problems.

The comparison of diagnoses derived from the interviews conducted independently with the mothers and children (Reich et al., 1982) produced similar results to the comparisons based on individual symptoms. Specifically, best agreement occurred with the diagnoses of antisocial personality disorder and enuresis with lower agreement for the category "possible depression" (p. 329).

The study by Herjanic and Reich (1982) is important for several reasons. First, it was one of the earliest studies to employ separate structured interviews with children and mothers in order to identify the level of agreement between the two informants. Second, care was taken to ensure that the sample size in the study was large enough to allow for meaningful statistical analyses. Finally, agreement was studied not only for individual symptoms but also for diagnostic categories. This made the results of the study highly relevant for the clinical assessment of children.

The study by Herjanic and Reich (1982) also highlights two limitations of earlier reports which described the level of agreement between children and parents. First, no information was provided about how many potential subjects refused to participate in the study. As a result, it is unclear whether the subjects who participated were truly representative of the

population from which they were drawn or whether they were biased in some way. Second, the subjects for the study were drawn from three somewhat different clinical services. As a result, it is plausible that the results of the study do not reflect the level of agreement which would be found amongst mothers and children living in the community. This is important, because as discussed earlier, it is possible that there are differences between the social and intrapersonal attributes of people with the same symptoms and disorders in clinic and community populations.

In 1986, Edelbrock et al. reported the results of a large study investigating the level of agreement between 299 children and their parents. The children were aged 6-18 years and had been referred to a psychiatric clinic. This was another of the group of studies conducted by Edelbrock and Costello which were referred to earlier (Costello et al., 1984; Edelbrock et al., 1985; Edelbrock et al., 1986; Edelbrock & Costello, 1988). This particular study focused on the size and direction of differences between reports obtained from parents and children, and the effect of child age on the level of agreement. The key findings from the study were consistent with those reported earlier by Herjanic and Reich (Herjanic & Reich, 1982; Reich et al., 1982):

1. Correlations between scores were higher on the scales describing behaviour or conduct disorders (mean r=0.42) than on those describing affective or neurotic disorders (mean r=0.19).
2. Parents reported more behaviour or conduct symptoms while children reported more affective or neurotic symptoms. The only exception to this pattern occurred in the area of

alcohol/drug abuse where children reported more problems than their parents.

3. The level of parent-child agreement was " ... strongly related to age [of the child]. Overall, parent-child agreement increased with age and was higher for children aged 14-18 (average r=0.35) than children aged 10-13 (average r=0.27) and 6-9 (average r=0.10)" (p. 186). The pattern of parent child agreement (See 2. above) was similar across all age groups.

In concluding, Edelbrock et al. (1986) emphasized the need for studies investigating the level of agreement between informants to "distinguish between *degree* of agreement between informants and *direction* of differences" (p. 189). For example, although parent and child scores in different symptom areas may show similar correlations, the magnitude of the differences between the scores and the direction of these differences may vary. Finally, Edelbrock et al. (1986) made the important point that in determining the meaning of differences between informants, further research was needed to identify the clinical significance of symptoms reported uniquely by one individual for the treatment and prognosis of child psychiatric disorders.

1.2122 Studies using Self-Report Measures

The studies by Kazdin et al. (Kazdin, French, & Unis, 1983; Kazdin, French, Unis, & Esveldt-Dawson, 1983; Kazdin, Esveldt-Dawson, Unis, & Rancurello 1983; Kazdin & Heidish, 1984) used a combination of different measures to collect information from children and parents. Kazdin, French, Unis, and Esveldt-Dawson (1983) adopted the approach of using

more than one measure for each symptom area because this allowed them to investigate whether the level of agreement between the informants was being influenced by any particular method. In addition, Kazdin, Esveldt-Dawson, Unis, and Rancurello (1983) included two different symptom areas, depression and aggression, in order to determine "whether correspondence of child and parent ratings varied as a function of these different symptom areas" (p. 402). Finally, information was collected from children, mothers, and fathers, which made it possible to study the level of agreement between children and fathers, in addition to that between children and mothers.

In their initial reports Kazdin and his colleagues (Kazdin, French, & Unis, 1983; Kazdin, French, Unis, & Esveldt-Dawson, 1983), focused on the level of agreement between children and parents with regard to the presence of childhood depressive symptoms. The first report described results obtained with 48 children (Kazdin, French, & Unis, 1983) and this sample was subsequently expanded to include a total 104 children, aged 5-13 years, all of whom were inpatients on a psychiatry ward (Kazdin, French, Unis, & Esveldt-Dawson, 1983). Like other investigators (Leon, Kendall, & Garber, 1980), Kazdin et al. (Kazdin, French, & Unis, 1983; Kazdin, French, Unis, & Esveldt-Dawson, 1983) found that the different measures of the children's depression were consistently correlated within a particular rater, however, mother-child and father-child correlations tended "to be low to moderate and generally not statistically significant" (Kazdin, French, Unis, & Esveldt-Dawson, 1983, p. 161).

In their third report, Kazdin, Esveldt-Dawson, Unis, and Rancurello (1983), focused on both childhood depression and aggression. For this study there were 120 children, aged 5-13 years, who again were all

inpatients on a psychiatric ward. The study employed two measures of depression and two measures of aggression. Once again, the investigators noted that "the most striking pattern in the [correlation] matrix is the high correlations for measures completed by the same rater. This suggests that the assessment method (rater) accounts for an important source of variance" (p. 408). Interestingly, a comparison of the size of the correlations between child and parent scores in the two symptom areas showed them to be similar (Depression, mean r=0.30, Aggression, mean r=0.31). Finally, the investigators noted that the correspondence between child and parent reports did not vary as a function of the assessment format (self-report or interview).

The mean depression and aggression scores reported by the children participating in the studies by Kazdin et al. ((Kazdin, French, & Unis, 1983; Kazdin, French, Unis, & Esveldt-Dawson, 1983), were significantly lower than those reported by the mothers. A similar pattern was evident with the fathers. This is one of the few studies in which children reported fewer internalizing symptoms than their parents and subsequent investigators have suggested that the result may be due to specific characteristics of the study sample. For example, Treiber and Mabe (1987) pointed out that parents in an inpatient setting may have less opportunity to observe their children directly and this may influence the level of correspondence between reports obtained from parents and children. Slotkin, Forehand, Fauber, McCombs, and Long (1988) suggested that in an inpatient sample, "children are less attuned to subjective feelings of distress or that other problems are interfering with their subjective appraisal" (p. 208). These comments highlight the possibility that the extent and nature of discrepancies between informants may vary

in different populations of children, that is they are "sample dependent".

1.213 Studies of Children with Psychiatrically Disordered Parents

In 1980, Weissman et al. described the level of agreement between 28 children, aged 6-17 years and their mothers, living in 12 families. Parents (not necessarily the mother) in the participating families " had at one time been subjects in a family-genetic study of affective disorders, but were rejected as probands in that study for diagnostic reasons" (p. 736). The study focused on agreement with regard to both the presence of symptoms of affective disorder in the child and the child's social functioning. Assessment of the children employed structured interviews and behaviour rating scales which were administered independently to the children and parents. Like Kazdin, Esveldt-Dawson, Unis, & Rancurello (1983), Weissman et al. (1980) found that the correlation between scores on the symptom and social adjustment scales was significant when scores were obtained from the same informant. In contrast, the correlation between scores was poor when scores were obtained from different informants.

Weissman et al. (1980) noted that the reports from mothers, but not those from children, differentiated children with diagnoses, regardless of whether the diagnoses were based on information obtained from the mothers or the children. In light of this, Weissman et al. (1980) concluded that when childhood problems are identified using self-report measures, mothers may be more sensitive reporters of their children's problems than the children themselves. However, two of the three self-report measures

completed by the children and mothers were different. This is important because it is plausible that the measures completed by the children were less sensitive to the presence of childhood diagnoses than those completed by the mothers. If so, this would provide an alternative explanation for the different diagnostic accuracy of the reports obtained from the mothers and children in this study.

In a second paper, Orvaschel, Weissman, Padian, and Lowe (1981) focused on the level of agreement with regard to the diagnoses derived independently from the interviews with the children and the mothers. Although the number of subjects was too small to make use of formal statistical analysis, Orvaschel et al. (1981) reported that there was only moderate agreement between the diagnoses derived from separate interviews with mothers and children. They suggested that the mothers were better able to provide factual information, however, information from the children was essential for knowledge about the children's internal state such as their moods, fears and feelings. Ivems and Rehm (1988) have recently noted that the same rater interviewed both the children and mothers in this study. In a similar study, Ivems and Rehm (1988) used a different person to interview children and mothers and found that when this was done there was a somewhat poorer level of agreement between the two informants. In light of this, Ivems and Rehm (1988) suggested that studies which use a single rater to conduct the interviews with both mothers and children may report better levels of agreement between the two informants.

Recently, Weissman et al. (1987) and Angold et al. (1987) have reported on the level of agreement amongst mothers and children from 104 families who had earlier participated in the Yale Family-Genetic Study of Major

Depression (Weissman et al., 1984). This study represented a continuation of the work described in the earlier reports by the same group (Weissman et al., 1980; Orvaschel et al., 1981). Fifty-six families in which there had been a depressed proband parent, and 35 families which had contained a normal control parent agreed to participate in the study. They represented 87.5% of the families who had earlier participated in the Yale Family-Genetic Study of Major Depression. The families contained a total of 220 children, aged 6-23 years, of which 125 children came from families with a depressed parent, and 95 children from the control families. Assessment of the children was by means of the Schedule for Affective Disorders and Schizophrenia for School-Aged Children, Epidemiological Version (K-SADS-E) (Orvaschel, Puig-Antich, Chambers, Tabrizi, & Johnson, 1982) which was administered independently to the children and parents (in all but six families this was the biological mother). The investigators noted that the study "comprises the largest sample of children, to date, in which the K-SADS-Epidemiological Version (K-SADS-E) interview and DSM-III criteria were used; the participation rate from the base sample is extremely high; a comparison group of children of normal nonpsychiatrically ill parents has been included; the interviewers were "blind" to the clinical status of the parents" (Weissman et al., 1987, p. 748). However, it is also important to note that the children in the study were considerably older than those in most similar studies with 20% of the participating children aged 19 years or more at the time of the study.

In their initial report, Weissman et al. (1987) described a number of important findings:

1. The overall agreement between children and mothers for most diagnoses was "statistically significant but modest with K [Kappa] values ranging from 0.07 to 0.36" (Weissman et al., 1987, p. 749). This pattern applied for children in both groups of families. Consistent with earlier studies, the best agreement was for conduct disorder.
2. The mother's reports were "primarily a subset of the children's reports" (Weissman et al., 1987, p. 749) and there was little evidence that mothers reported another diagnosis when depression was reported by the child.
3. The strongest effect of child's age occurred with major depression where the level of agreement between children aged greater than 19 years and their mothers was much worse than between mothers and younger children. This was due to the older children, particularly the females, reporting higher rates of depression, a pattern which was not reflected in the reports from mothers.
4. If the parent was currently depressed, "there was a substantial improvement in agreement on [the] presence of any diagnosis in the child (K=0.74) but not for a specific diagnosis. This indicates that parents who were experiencing a major depression compared with those who were not were more sensitive to their children's psychopathology, even though they did not agree with their children about the symptom manifestations of the child's disturbance" (p. 750).
5. A variety of other factors, including the sex and marital status of the parent, the child's IQ, treatment history and living arrangements (living with parents or elsewhere), did

not have a major impact on the level of agreement between children and parents.

In their second report, the investigators (Angold et al., 1987) focused specifically on agreement between parent and child reports of childhood depressive symptoms. Angold et al. (1987) noted that "positive parental reports of dysphoria in the absence of a positive child report are distinctly uncommon (occurring in only 6.3% of cases). Thus the children whose parents say that they have been dysphoric are mostly a subgroup of those who report themselves to have been dysphoric" (p. 904). Interestingly, Angold et al. (1987) found that the depressed proband parents reported significantly more symptoms of dysphoria in their children than did the control parents. In contrast, there was no significant difference in the number of dysphoric symptoms reported by the children. Although it was not clear whether the proband parents were depressed at the time they reported on their child, the result suggested the possibility that parental psychopathology may have influenced reports from the parents, while having little or no effect on the children's reports. However, Angold et al. (1987) cautioned that in subsequent analyses (Weissman et al., report in preparation, cited in Angold et al., 1987) there were differences in the number of depressive diagnoses identified from the children's reports.

The possibility that factors, such as maternal psychopathology, may influence specifically reports obtained from one informant but not from others is an important issue. In the past many studies have relied heavily on reports from mothers. It has been assumed that when a factor "external" to children's behaviour, showed a significant association with mothers' reports of children's behaviour, the association reflected a

relationship between the external factor and the children's behaviour. What has received less attention is the possibility that the primary relationship is one between the external factor and mothers' perceptions of their children's behaviour. For example, if self-reported maternal psychopathology showed a significant association with maternal reports of childhood behaviour problems, it could be assumed that there is a relationship between maternal psychopathology and childhood behaviour problems. However, the nature of this relationship is less clear if no significant association is found when reports of children's behaviour are obtained from the children or their fathers. This finding would suggest that the relationship identified from the mother's reports is "informant specific". That is, there is a significant relationship between this factor and childhood emotional and behavioural problems only with reports from a specific informant.

Finally, Kashani, Orvaschel, Burk, and Reid (1985) employed the Diagnostic Interview for Children and Adolescents (DICA & DICA-P) (Hodges, McKnew, Cytryn, Stern, & Kline, 1982) to study the level of agreement between children and parents. The study focused on agreement between 24 adult patients with a diagnosis of major affective disorder, and their 50 children, aged 7-17 years. Results from this study were consistent with the studies reported earlier by Herjanic and Reich (Herjanic & Reich, 1982; Reich et al., 1982). In particular, depression was found to be the most common disorder when diagnosis was based on children's reports of their own behaviour and attention deficit disorder and oppositional disorder were the most frequent disorders when diagnosis was based on parental report. Finally, diagnostic agreement between parents and children was reported as "extremely low for all Axis I disorders examined" (p. 439).

Kashani et al. (1985) offered two interesting explanations for their findings. First, they suggested that parents and children "may use different criteria of severity for determining when a behavior is worth mentioning" (p. 440). For example, although a parent may be aware that their child worries, they may not consider this a problem and may not report it during their interview. Children, on the other hand, may consider that their worrying is a problem and thus report it during their interview. Second, Kashani et al. (1985) noted that in their study "a majority of the children diagnosed as depressed (on the basis of children's reports) were considered by parents to have attention deficit or oppositional disorder" (p. 440). They suggested that parents may have been attending to the attentional difficulties and acting out behaviors, which coincided with depression for some children, while ignoring the child's affective problem, "Therefore, the child's experience of depression may be interpreted by parents to be a behavior problem and not an affective disturbance" (p. 441). Finally, Kashani et al. (1985) highlighted three potential sources of variance in the assessment of childhood emotional and behavioural disorders: criteria variance, information variance and informant variance. They suggested that the use of specified criteria, such as the DSM-III had improved reliability by reducing information variance and the use of new assessment techniques, such as structured interviews, had reduced information variance. What remained as a problem was informant variance, "when more than one informant is used in diagnostic evaluation, the sine qua non result is disagreement" (p. 437).

1.214. Studies of Children living in the General Community

In 1987 two reports were published which described the level of agreement between children and parents living in the community. One report described the results of semi-structured interviews with children and parents (Verhulst et al., 1987) while the second employed various rating scales to compare reports describing children living in military families in the USA (Jensen, Traylor, Xenakis, & Davis, 1988; Jensen, Xenakis, Davis, & Degroot, 1988). These were substantive studies focusing specifically on agreement about children's behaviour for children living in the general community.

The study by Verhulst et al. (1987) employed a semi-structured interview rather than one of the more highly structured interview formats employed in earlier studies. Verhulst et al. (1987) adopted this approach because they reasoned that the lengthy, somewhat rigid nature of more structured interviews may have affected the level of agreement found between reports from children and parents. They suggested that this could occur because more structured interview formats may have limited the rapport which could be achieved with the children. As well, structured interview formats provide limited opportunities for observing children's behaviour, mental status and cognitive skills.

The children who participated in the study by Verhulst et al. (1987) had earlier participated in a large epidemiological study investigating the prevalence of emotional and behavioural problems in 8 and 11 year old Dutch children. From this larger sample, 78 children were selected for semi-structured interview on the basis that they scored at or above the

85th percentile on either the Teacher Report Form (TRF) (Achenbach & Edelbrock, 1986) or the Child Behavior Checklist (CBCL) (Achenbach & Edelbrock, 1983). A comparison group of 75 children were randomly drawn from the remainder of the sample. Interviews were obtained with 116 of the children and their parents, representing a response rate of 76%. The interviewed parents included mothers (58%), fathers (7%) and both parents jointly (35%). The semi-structured interviews employed with the children and parents were different. Interviews with the children employed the Child Assessment Schedule (Hodges et al., 1982) while those with parents used a modified version of the semi-structured interview developed by Graham and Rutter (1968).

The results of the study by Verhulst et al. (1987) varied from those of previous similar investigations (Edelbrock et al., 1985; Weissman et al., 1987). In particular, the correlation between the child and parent total problem scores of r=0.58 was considerably higher than the correlation of r=0.17 between child and parent total problem scores reported in the study by Edelbrock et al. (1986). As well, in contrast to the studies by Edelbrock et al. (1985) and Weissman et al. (1987), there was no association between the child's age and the level of agreement between children and parents. Finally in most areas, the parents reported more problems than the children.

Trying to identify the reasons for the different results in the studies by Verhulst et al. (1987), Edelbrock et al. (1985, 1986), and Weissman et al. (1987) highlights the difficulties inherent in comparing results from studies which have investigated the level of agreement between parents and children. In the three studies the samples of children were drawn from three different populations, the ages of the children varied

greatly, the parent interviews were conducted with individual parents in the studies by Edelbrock et al. (1985, 1986) and Weissman et al. (1987), but in many cases involved two parents in the study by Verhulst et al. (1987), and the interview formats were different in each of the three studies. The variation across the studies in so many key areas makes it difficult to know which factors may have contributed to the different results obtained in the studies.

The study by Jensen, Traylor, Xenakis, and Davis (1988) which was described earlier, focused on the level of agreement between a group of 108 children, aged 6-11 years, and their parents. The study focused particularly on inter-parent and parent-teacher agreement, however, the correlations between child and parent scores were also reported. These ranged from r=0.02 to r=0.45 and the authors noted that the correlations between mothers and sons, in particular, were somewhat higher than previous reports of correlations between parent and child scores. They suggested that this "may reflect the effects of separately reporting the parent and child correlations by both gender of the parent and gender of the child" (p. 447). The correlations between mothers and sons were also found to be significantly higher than the correlations between scores from fathers and sons. This finding was consistent with that of Ivens and Rehm (1988) who also reported that the level of agreement between mothers and children was slightly higher than that of fathers and children.

In order to determine whether the psychological adjustment of the parents affected the level of agreement between child and parent scores, correlations were computed between the difference between parent and child scores and the parent's HSCL score after all the scores had been

subjected to z transformation (Jensen, Xenakis, Davis, & Degroot, 1988). The results of these analyses varied depending on the sex of the parent and child. The authors concluded that, in general, there were "consistent differences in mothers' and children's perceptions of child symptoms as a function of mothers' own psychiatric symptoms .... these effects appear to be greater for sons than daughters" (p. 449). Father-child differences were reported as "approximately equally correlated with fathers' symptoms for both boys and girls" (p. 449). The same methodology was used (Jensen, Xenakis, Davis, & Degroot, 1988) to examine the effect of a wide range of other variables on the level of agreement between parents and children. These variables included the child's age, the parent's familiarity with the child, the family's socioeconomic status, birth order, the family's size, the child's tendency to respond in a socially desirable fashion, the nature of the symptoms being compared (e.g. externalizing versus internalizing), family stress levels, and the child's friendship patterns. In general, none of these variables showed a strong association with the size of the difference between the child and parent score.

Jensen, Xenakis, Davis and Degroot (1988) cautioned that the subjects in their study all came from military families and this may limit the extent to which their findings can be generalized to other populations. They also made the important point that the results may not apply to families in which a child has been referred to a mental health clinic. Finally, it is important to note that the measures completed by the children and parents in this study were different. The measures completed by the children focused specifically on depression and anxiety while those completed by parents had a more general focus.

1.215 Studies which included Children living in the General Community and Children referred to Mental Health Clinics

One of the very few studies which compared reports from children and parents in both community and clinic samples is that by Mokros, Poznanski, Grossman and Freeman (1987). The children were aged 6-12 years, and there were 110 children in the community sample and 34 children in the clinic sample. Children and parents in each group were separately interviewed by clinicians using the Children's Depression Rating Scale-Revised (Poznanski, Grossman, Buchsbaum, Banegas, Freeman, & Gibbons, 1984).

A key finding from the study by Mokros et al. (1987) was that the level of agreement between children and parents appeared to be "sample-dependent" (p. 621). Specifically, children in the community tended to report more symptoms than their parents while this situation was reversed in the clinic sample where the parents reported more symptoms than the children. Mokros et al. (1987) noted that "In contrast with clinically referred children, children from a non-clinical population exhibited a greater tendency to rate themselves more severely than their parents on a number of specific symptoms including behavioral as well as ideational symptoms" (p. 621). This is an important result which highlights the importance of not assuming that the attributes of clinic and community samples are the same. Instead, the result suggests that there may be differences in the pattern of reports from children and parents describing childhood emotional and behavioural problems in clinic and community populations.

The results of the study by Mokros et al. (1987) need to be interpreted with caution because of limitations in the design of the study. In particular, the children in the community all attended the same school and the children in the clinic sample had all been referred to a specialized Affective Disorders Clinic. As well, the "parents" differed somewhat in the two samples. Ninety-three percent of the community parents were mothers while the clinic parents included a more heterogeneous group of mothers, fathers, grandmothers and in some cases two parents who appear to have been jointly interviewed. Finally, the children were all relatively young and no information was provided about the response rate in either the clinic or community samples.

1.22 STUDIES OF AGREEMENT BETWEEN MOTHERS AND FATHERS

This section reviews studies which have investigated the level of agreement between mothers and fathers when describing emotional and behavioural problems in their child.

In the early 1960s, several studies compared reports of childhood behaviour problems obtained from mothers and fathers (Becker, 1960; Peterson, Becker, Shoemaker, Luria, & Hellmer, 1961; Hafner et al., 1964; Miller, 1964). For example, Hafner et al. (1964) compared reports of childhood anxiety symptoms obtained from parents and children. Hafner et al. (1964) noted that correlations between the parent scores ($r=0.53$ to $r=0.55$) were consistently higher than those between the child and parent scores ($r=0.36$ to $r=0.50$).

In an interesting study, Miller (1964) reported on the agreement between parents of 36 children from 18 two-parent families "applying to the Louisville Child Guidance Clinic, who had at least two children of the same sex between six and ten, one whom they described as disturbed (the referred child) and one about whom they felt comfortable (the nonreferred child)." (p. 71). The study employed a Q sort procedure composed of 80 statements covering various dimensions of personality and interpersonal dimensions. The average correlation between parents was higher for the nonreferred children (r=0.68) than for the referred children (r=0.53). Miller (1964) pointed out that the poorer agreement about the behaviour of the referred children is of particular concern because it is these children that clinicians are seeing in mental health clinics for assessment.

Ferguson et al. (1974) described inter-parent agreement for 112 children referred to a psychology clinic and 105 children from the same geographical area who had not been referred to a clinic. The children were aged 5-11 years and parents described their children's behaviour using a behaviour checklist developed by Ferguson et al. (1974) The checklist included "154 items referring to interpersonal and symptomatic behaviours" (p. 171) and items were scored by first checking all items that "apply at all" to the child and then checking items which were "characteristic" of the child. The study focused on comparing parent perceptions of referred and non-referred children, however, inter-parent agreement was also reported for children where both parents completed a checklist (approximately 72% of the referred families and 91% of the non-referred families). Ferguson et al. (1974) used factor analyses to compare the reports obtained from the four sets of parents (clinic mothers, clinic fathers, non-clinic mothers, and non-clinic fathers).

These analyses identified somewhat different factors for each set of parents and Ferguson et al. (1974) concluded that the results "suggest that mothers and fathers of normal and clinic-referred children are sensitive to somewhat different behavioral dimensions" (p. 179).

In order to identify the level of agreement between parents, Ferguson et al. (1974) derived an index of interparent agreement by "dividing those items which were checked in either the characteristic or applicable column by either parent into twice the number of characteristic or applicable items that parents agreed on." (p. 173). Analyses of variance were then used to compare the level of agreement between parents in the clinic and non-clinic groups. In contrast to the findings of Miller (1964), Ferguson et al. (1974) reported that there was no significant difference in the level of interparent agreement between the normal and clinic-referred children. It is possible that the different results reflect the use of different measures and different statistics in the two studies.

Jacob et al. (1982) used five different statistics in a study of interparent agreement. Subjects in the study included fifty-two index families in which there was a delinquent male child, aged 10-18 years, who were identified through community-based probation offices. Six additional index families were recruited through local schools. The remaining 38 control families were described as "normal" but the means by which they were recruited was not reported. All parents completed the Behavior Problem Checklist (Quay, 1977) which consists of 54 items and has four scales labelled "Conduct Problems", "Personality Problems", "Inadequacy-Immaturity", and "Socialized Delinquency".

The consistency between reports from mothers and fathers in the study by Jacob et al. (1982) was investigated using several different approaches. First, correlation coefficients were calculated and these revealed that the correlation between parent scores in the index families was smaller (r=0.18) than that in the normal families (r=0.61). For both groups of families, the correlations were larger between parent scores rating conduct problems than between scores rating personality problems. Second, Percentage Agreement and Effective Percentage Agreement (defined as the ratio of the number of agreements to the total of all observations where one or both parents acknowledged occurrence) were used to compare reports from the two sets of parents. Results obtained using the Percentage Agreement Index suggested substantial interparent agreement as to the presence of specific items (69% in index families and 78% in control families), while use of the more stringent Effective Percentage Agreement suggested lesser agreement (53% in index families and 34% in control families). Interestingly, when the kappa statistic (which corrects for chance agreement) was used to measure interparent agreement, the apparent differences between the index and control groups were largely eliminated. This suggested that the differences identified by the percentage agreement indices were largely artefacts arising from the inability of the indices to correct for the differing levels of chance agreement in the two groups.

Two approaches were used in the study by Jacob et al. (1982) to investigate the level of agreement about the number of the children's problems reported by mothers and fathers. First, the significance of the difference in the scores reported by mothers and fathers was tested using t tests. These analyses revealed that fathers in the control families endorsed significantly more items than the mothers. A similar pattern

was evident in the index families although the differences did not achieve statistical significance. Second, the absolute difference scores (defined as the absolute difference between parent scores on relevant scales) in the index and control families were compared. These comparisons showed that the absolute differences for the total problem score and the Personality Problem scores were significantly greater in the index families than the control families.

Both Jacob et al. (1982) and Jensen, Traylor, Xenakis, and Davis (1988) have suggested that further studies investigating interparent agreement are needed. For example, Jacob et al. (1982) pointed out that "the study of parent agreement, its correlates and consequences, has barely gone beyond the preliminary stages of investigation" (p. 606) and they suggested that their own findings needed replication with a larger more representative sample of male and female children. Jensen, Traylor, Xenakis, and Davis (1988) noted that although stronger correlations have been reported between interparent reports, than between child and parent reports, "undue confidence in interparent reliability may not be well founded" (p. 442). They pointed out that most studies which have reported interparent reliability have done so only incidently in the context of validating behaviour checklists. As well, most studies have reported only correlations between scores and ignored the possibility that despite good correlations, there may be large differences in the actual number of problems reported by mothers and fathers. Jensen, Traylor, Xenakis, and Davis (1988) concluded "The assumption that either parent can be used as the reporter in studies of child psychopathology in community-based samples does not appear to rest on any firm empirical data" (p. 442).

1.23 STUDIES OF AGREEMENT BETWEEN PARENTS AND TEACHERS

This section reviews studies which have investigated the level of agreement between parents and teachers when describing emotional and behavioural problems in children.

In contrast to the studies focusing on agreement between children and parents, which have largely focused on clinic-referred children or the offspring of psychiatrically disturbed parents, studies of agreement between parents and teachers have largely focused on children in the general community.

In the epidemiological study of Berkshire children referred to earlier, Mitchell and Shepherd (1966) compared the number of problems reported by parents and teachers. For the purpose of these analyses, children were assigned to one of four groups, according to the number of problems they were reported as having by their parents. A similar grouping was undertaken on the basis of the number of problems reported by teachers. A chi-square statistic was then used to determine whether there was a significant association between the number of problems reported by parents and teachers. This revealed a significant ($p<0.001$) association between the number of problems reported by each informant, however, the investigators also noted that "more than a third of the children who were reported free of deviant behaviour at home, nevertheless, exhibited at least one problem in school and that nearly half of the children whose parents underlined seven or more deviant traits were apparently problem-free at school." (p. 251). When a comparison was based on the 10% of children exhibiting the greatest number of problems, "only about one

child in five picked as in the worst 10 per cent. by one questionnaire was also picked by the other" (p. 253).

As part of a comprehensive epidemiological study of childhood problems conducted on the Isle of Wight (Rutter, Tizard, & Whitmore, 1981), parents and teachers of 2193 children, aged 10-11 years, were asked to complete behaviour questionnaires describing the children's problems (Graham, 1967). Two hundred and eighty-four children who scored above established cutoff scores on the questionnaires were selected for more intensive investigation which included parental interview and psychological testing of the child. This approach made it possible to study agreement between parent and teacher reports at several levels including:

1. The number of children identified by each informant as scoring above the cutoff scores on the questionnaires used in the study.
2. The type of problems identified by parents and teachers on the questionnaires.
3. The type of problems identified "spontaneously" (Graham, 1967, p. 31) by parents and teachers when describing the children.

Consistent with the results found in the study by Mitchell and Shepherd (1966), there was little overlap between the children identified as having problems by parents and teachers. For example, while the percentage of children scoring above the cutoff scores on the parent and teacher questionnaires were similar (6.0% on the parent questionnaire and 7.1% on the teacher questionnaire), only 0.7% of these children scored above the cutoffs on both questionnaires. In commenting on this finding,

Graham (1967) noted "By chance about half this number would have been expected, so although there was somewhat greater agreement than might have been expected by chance alone, in general the questionnaires selected different children." (p. 30). Interestingly, children picked out on the basis of results from teacher or parent questionnaires were equally likely to be rated as "abnormal" (p. 30) on the basis of a more detailed assessment by a psychiatrist. "Psychiatrists therefore agreed equally well with both parents and teachers in this respect" (p. 30). The type of problems (neurotic or antisocial) identified by parents and teachers did not differ greatly although it appeared that teachers may have selected more antisocial boys.

Graham (1967) suggested a number of plausible reasons for the differences between the reports obtained from parents and teachers. These included the possibility that children behave differently at home and at school or alternatively, "the symptoms of disturbance that children show at home differ from those shown in school. Bedwetting is an obvious example but others are less obvious" (p. 31). As well, Graham (1967) suggested that some parents appeared to be denying problems which they knew to be present in their child and this influenced the level of agreement between parent and teacher reports. Several other factors were investigated including the child's I.Q., the occupational class of the parents, and "broken homes" (p. 31), however, these did not appear to be associated with the level of agreement between parents and teachers. Finally, Graham (1967) noted that the spontaneous complaints about children made by parents and teachers differed markedly with parents describing more neurotic symptoms (crying, worrying and unusual fears), temper tantrums and irritability. Graham (1967) suggested that the lack of opportunity for expressing these symptoms in the classroom, and the greater

likelihood that parents see their children at the end of the day when children who are tired and hungry are more likely to express these feelings, may in part account for these differences.

Since the Isle of Wight study there have been a number of smaller studies which have compared parent and teacher reports of children's behaviour. The studies have focused on samples of children selected from schools (Touliatos and Lindholm, 1981), the general community (Garrison & Earls, 1985; Anderson et al., 1987; Fergusson & Horwood, 1987a, 1987b, 1989), and mental health clinics (Schaughency & Lahey, 1985; Achenbach & Edelbrock, 1986).

Touliatos and Lindholm (1981) compared parent and teacher ratings for 1008 children, enroled in kindergarten to eighth grade, using the Behavior Problem Checklist (Quay, 1977). Both mothers and fathers reported significantly more problems than teachers on the scales rating Conduct Problem, Personality Problem, and Inadequacy-Immaturity. Correlations between the parent and teacher scores were described as "low or low to moderate" (p. 352) and ranged from r=0.06 for the correlation between mothers and teachers on the scale rating Psychotic Signs to r=0.45 for the correlation between mothers and teachers on the Conduct Problem Scale. Correlations tended to be higher for boys than girls. Touliatos and Lindolm (1981) suggested that the greater number of problems reported by parents may be due to the parents being "less accepting and more sensitive to children's difficulties because they have more invested in their offspring, that parents expect more and are less tolerant of children's maladaptive behavior possibly because they do not have as much formal knowledge about child development as teachers, or that parents and teachers see children in different circumstances and

situations and that variations in their perceptions of disorders have a basis in reality." (p. 352).

Several other studies have focused on agreement between parent and teacher reports for groups of children selected from the general community (Garrison & Earls, 1985; Anderson et al., 1987; Fergusson & Horwood, 1987a, 1987b, 1989; Bird, et al., 1988; Verhulst & Akkerhuis, 1989). For example, Garrison and Earls (1985) reported on a pilot study of 62 children, aged 6-7 years drawn from a larger birth cohort of young children. The children's parents completed the Child Behaviour Checklist (Achenbach & Edelbrock, 1983) and their teacher completed the Teacher Report Form (Achenbach & Edelbrock, 1986). Consistent with many other studies (Achenbach et al., 1987), the correlation between scores obtained from parents in the study ($r=0.52$) was considerably larger than that found between mothers and teachers ($r=0.14$) or between fathers and teachers ($r=0.13$). Once again, the teachers generally reported fewer problems than the parents and Garrison and Earls (1985) suggested that the teachers tended to focus on "behavior symptoms as they related to the academic and compliance dimensions of the school context and they were generally reluctant or unable to evaluate symptoms outside of those main areas." (p. 78).

The study by Bird et al. (1988) focused on the prevalence of childhood "maladjustment" in the community in Puerto Rico. The study employed the recommended cutoff scores on the Child Behavior Checklist (CBCL) (Achenbach & Edelbrock, 1983) and the Teacher Report Form (TRF) (Achenbach & Edelbrock, 1986) to identify a group of children who were then subjected to a more detailed assessment procedure. The proportion of children identified on the CBCL was 37.8% and on the TRF was 23.3%.

This represented a much higher proportion of children than were identified by parents and teachers in the first stage of the Isle of Wight study. In discussing this finding, Bird et al. (1988) suggested that higher cutoff points may be needed to improve the specificity of the CBCL, at least with Puerto Rican children. The results of a recent study by Sawyer, Sarris, Baghurst, Cornish and Kalucy (in press) suggest that similar adjustments may be required with studies of Australian children. In the study by Bird et al. (1988) the majority of children were classified as cases by only the parent or the teacher and " ... discrepancies in classification between parents and teachers occurred in 32.5% of the subjects, with a 14.3% agreement on the positive and 53.2% agreement on the negative screen" (p. 1123).

The study by Verhulst and Akkerhuis (1989) which was described earlier, employed the CBCL and TRF to compare parent and teacher reports of a randomly selected group of Dutch children, aged 4-12 years. The study reported the mean number of problems described by each informant and the size of the correlations between the scores for the 94 individual problem items common to each checklist. Consistent with earlier studies, the majority of the differences between the mean scores occurred because parents reported more problems than teachers. The average correlation between parent and teacher total problem scores was r=0.27 for the younger children (4-5 years) and somewhat higher, r=0.35, for the older children (6-12 years). There was no significant difference in the size of the correlation between parent and teacher scores for the male and female children.

The reports by Fergusson and Horwood (1987a, 1987b, 1989) are important because they draw attention to a theoretical model which was "designed to

estimate the trait and method specific components of maternal and teacher ratings" (1987a, p. 249). The reports focused on conduct problems in a birth cohort of 1103 New Zealand children. Maternal and teacher ratings were obtained when the children were aged 7, 8, and 9 years, and ratings from the children when they were aged 9 years. The maternal and teacher reports consisted of versions of the Rutter (Rutter et al., 1981) and Conners questionnaires (Conners, 1969; Conners, 1970). Reports from the children were obtained by means of a child self-report scale (Beitchman, Raman, Carlson, Clegg, & Kruidenier 1985).

Fergusson and Horwood (1987a) highlighted three potential sources of variation in mother and teacher ratings of childhood conduct disorder: "(a) variation due to true trait effects; (b) variation due to method specific factors; (c) variation due to random errors of measurement" (p. 250). They suggested that a substantial proportion of the variation in the mother and teacher scores may not be due to trait variation but instead arises from method effects. In order to investigate this possibility, Fergusson and Horwood (1987a, 1987b) employed a linear structural equation model. In their first two reports (Fergusson & Horwood, 1987a, 1987b) they used the model to investigate the characteristics of mother and teacher ratings, and in their final report (Fergusson & Horwood, 1989) they incorporated the ratings obtained from the children into their structural equation model.

Fergusson and Horwood (1987a) reported that their analyses showed that "a considerable component of variation in teacher and maternal scores reflects the effects of method factors rather than trait variance." (p. 256) and "the apparent correlations between maternal depression, family life events and maternal ratings of conduct disorder arose primarily from

the effects of method factors rather than because of an association between these variables and the child's tendencies to conduct disorder. It seems likely that independently of the child's behavioural tendencies, mothers who are depressed or under stress may see their children as being more troublesome and prone to conduct disorder (or that children with mothers who are depressed or stressed tend to develop situation specific conduct disorders)." (Fergusson & Horwood, 1987b, p. 269). In the final paper in this series, Fergusson and Horwood (1989) reported that only approximately 28-40% of the variation in parent and teacher ratings was due to variation in the "child's generalized (i.e. trait) behavioural tendencies" (p. 374).

The reports by Fergusson and Horwood (1987a, 1987b, 1989) are important and need replication. Verhulst and Akkerhuis (1989) argued, on the basis of the high test-retest reliability of the measures used in their study and the correction of their correlations for attenuation, that their results were unlikely to have been limited much by the unreliability of their measurement procedures. Despite this, they noted that the correlations between parent and teacher scores in their study were still modest and acknowledged that "we regard method-specific factors of ratings from different informants as of empirical interest and, we would like to add, of clinical interest as well" (Verhulst & Akkerhuis, 1989, p. 134).

1.24 STUDIES OF AGREEMENT BETWEEN CHILDREN AND TEACHERS

This section reviews studies which have investigated the level of agreement between children and teachers when describing emotional and

behavioural problems in the child. In contrast to the large number of studies investigating the level of agreement between children and parents, there have been relatively few studies which have investigated the level of agreement between children and teachers.

Studies which have obtained reports from both children and teachers (Cowen et al., 1965; Ledingham, Younger, Schwartzman, & Bergeron, 1982; Reynolds, Anderson, & Bartell, 1985) have generally found a relatively low level of agreement between these informants. For example, Ledingham et al. (1982) compared reports from 801 children in primary school with reports obtained from their teachers using the Pupil Evaluation Inventory (Pekarik, Prinz, Liebert, Weintraub, & Neale, 1976). Correlations between children and teachers were low to moderate, ranging from r=0.11 to r=0.48. There was no evidence that correlations increased in higher grades, however, correlations were higher for ratings of aggression than for withdrawal and likability. Ledingham et al. (1982) suggested that social desirability appeared to have a significant influence on the children's self-reports. They also suggested that ratings by teachers and children may reflect the nature of the interactions between the teacher and the child. This is an interesting concept which may also explain some of the differences in reports obtained from fathers and mothers. For example, if a child has a much poorer relationship with one parent than the other, it is possible that the former parent will identify the child as having more problems.

One of the largest studies which has reported on the level of agreement between children and teachers was that by MacMillan, Kolvin, Garside, Nicol, and Leitch (1980). The primary focus of the study was on the effectiveness of a school-based intervention program for children with

psychiatric disorders. However, in the course of the study reports about children's behaviour were obtained from children and their teachers. The study included 1723 children, aged 10-12 years who were attending six "junior high or comprehensive schools" (p. 266). All children in the study completed the Junior Eysenck Personality Inventory (JEPI) (Eysenck, 1965) and teachers completed the Rutter Teacher Scale (Rutter et al., 1981).

The level of agreement between the children and teachers was reported both in terms of the degree of overlap of cases identified using the recommended cutoff scores on the two questionnaires, and the size of the correlations between scores on the questionnaires. The overlap between cases selected by the JEPI Neuroticism Scale and by the Rutter Teacher Scale was only 18% on the Rutter Neurotic Scale, and 15% on the Rutter Antisocial Scale. Consistent with these results, the correlations between scores on the JEPI Neuroticism Scale and the Rutter Scales were virtually zero (Rutter Total Score, r=-0.03; Rutter Neurotic Score, r=0.02; Rutter Antisocial Score, r=-0.03).

MacMillan et al. (1980) cautioned that the questionnaires used by the children and teachers in their study had "little content in common" (p. 275) and this may explain some of the discrepancies between the child and teacher reports. However, they pointed out that the difference between the child and teacher reports also may have occurred because the teacher ratings focused primarily on behaviour at school, while reports from the children likely covered a variety of settings. This is an important issue which has received little attention and which may also explain, in part, some of the discrepancies between reports from children and parents. For example, while the report from a parent likely reflects the

parent's perception of his or her child's behaviour in the context of the parent's relationship with the child, no similar limitation applies to the report from the child. It is plausible that reports from children reflect their perception of their behaviour in a wide range of settings including at home, while at school, with peers, alone by themselves, and with parents and teachers.

1.3 THE SIGNIFICANCE OF DISCREPANCIES BETWEEN INFORMANT REPORTS FOR EPIDEMIOLOGICAL STUDIES INVESTIGATING THE PREVALENCE OF CHILDHOOD EMOTIONAL AND BEHAVIOURAL DISORDERS

This section discusses the significance for epidemiological studies of discrepancies between reports from parents, teachers, and children.

Studies investigating the prevalence of childhood emotional and behavioural problems have generally employed a categorical approach when describing childhood problems. Recently, Lobovits and Handal (1985) drew attention to the potentially important effect that discrepancies between informants may have on such estimates. Lobovits and Handal (1985) pointed out that many studies failed to describe how information from different informants was integrated to form a single prevalence estimate, "As a result it is unclear whether such discrepancies may differentially effect reported prevalence rates, and if so, in what manner" (p. 47). In their study of 50 children, aged 8-12, attending a psychology clinic, Lobovits and Handel (1985) reported that the prevalence of childhood depression was 34% when based on interviews with the child and only 22% when based on information obtained from parents.

The problem discrepancies create for epidemiological studies was also highlighted in a report by Hulbert et al., (1986). This report described interparent agreement for 360 clinic-referred children, aged 3-18 years. Both parents independently completed the Personality Inventory for Children (PIC) (Wirt, Lachar, Klinedinst, & Seat, 1984). The study then compared the level of agreement between the parents about whether or not their child scored above the recommended cutoff scores on the PIC. The cutoff scores were developed by Wirt et al. (1984) to facilitate differentiation of normal children from those who are likely to have clinically significant problems.

Hulbert et al. (1986) pointed out that it had been assumed that reports from mothers were sufficient to identify children who scored in the clinical range and, as a result, separate reports from fathers were rarely obtained. In their study, the percentage of parents who agreed that their child scored above the cutoff score on the Personality Inventory Scales "spanned a rather narrow range, from a low of 67% to a high of 81%" (p. 121). Hulbert et al. (1986) suggested that these results represented a relatively low level of agreement between mothers and fathers. Interestingly, they pointed out that the results of these comparisons corresponded poorly to the results of comparisons which employed correlation coefficients to identify the level of agreement between informants.

The problem discrepancies between informants cause for epidemiological studies was further highlighted in the study by Anderson et al. (1987). The subjects in the study were 792 children, aged 11 years, who were participating in a longitudinal study of the "health, development, and behavior of a large representative sample of New Zealand children" (p.

70). In their study, Anderson et al. (1987) focused on the prevalence and the degree of overlap of DSM-III disorders in the children, using the children as the major source of information. Information was obtained from the children by means of DISC interviews. Parents completed the Rutter A Scale (Rutter et al., 1981) and "two additional sets of behavioral items relating to attention deficit disorder (ADD) and depression" (p. 70) and information from teachers was obtained by means of the Rutter B Scale (Rutter et al., 1981).

In order to try to integrate information from the different informants and produce prevalence estimates, Anderson et al. (1987) adopted the approach of defining levels of certainty for case identification. The levels were defined on the basis of the number of information sources who reported the presence of relevant symptoms and ranged from "Level 1", where diagnostic criteria were met independently by more than one source, to "Level 4", where diagnostic criteria could only be met by combining symptoms from all three sources (parents, children, and teachers). Interestingly, the number of children meeting the criteria for a Level 1 or Level 2 diagnosis was 7.3% but an additional 4.2% of the children met the criteria for a Level 4 diagnosis.

1.4 SUMMARY

Discrepancies between reports obtained from children, parents, and teachers are an important problem hindering both research, and the accurate clinical assessment and treatment of childhood emotional and behavioural disorders. Currently there are no guidelines available to advise clinicians or researchers as to how they should weight discrepant

reports in order to evaluate accurately childhood disorders. Thus, assessments made by clinicians and researchers will vary, depending on the choice of informants, the extent of agreement between the informants, and finally, whether information from one informant is weighted more heavily when disagreement exists.

Commenting on the problem of discrepant reporting, Herjanic and Reich (1982) have pointed out that it cannot be assumed that parental reports are more accurate than those obtained from their child simply because parents are older. Furthermore, both Hill (1985) and Herjanic and Reich (1982) have made the obvious point that children may be the only, or at least the most important, source of information about subjective distress involving feelings such as guilt, sadness, anxiety, or pain. Hill (1985) pointed out that, although personal distress may be inferred from public behaviour, parent reports may minimise the child's subjective suffering because of ignorance or vested interest. In addition, emotional turmoil in the parents may affect their capacity to identify accurately their child's distress (Herjanic & Reich, 1982).

An important limitation of many previous studies which have investigated the level of agreement between different informants is that they used different measures to obtain reports from different informants. This makes it difficult to know whether the differences between reports were due to the different measures used or the different characteristics of the informants. As well, most studies focused on reports from only two informants which makes it difficult to obtain a clear picture of the differences which may exist between reports from parents, teachers and children. A further limitation of earlier studies is that they focused either on children who had been referred to mental health clinics

(Herjanic & Reich, 1982; Edelbrock et al., 1986) or children living in the general community (Jensen, Traylor, Xenakis, & Davis, 1988; Verhulst et al., 1987). Few studies have included samples of both clinic-referred children and children living in the general community. This makes it difficult to know the extent to which results obtained in previous studies apply only to certain populations of children.

Finally, there have been few studies which have attempted to identify factors which may be associated with the poor agreement between reports from different informants and results to date are contradictory. For example, while Moretti et al. (1985) reported that parent psychopathology may bias parent observations of their children, Weissman et al. (1987) reported better agreement between parents and children when the parent was depressed when they described their child's problems.

1.6 THE AIMS OF THIS STUDY

The primary aim of this study was to describe the level of agreement between children, parents, and teachers in the general community when reporting on emotional and behavioural problems in the child. In addition, the study investigated reasons for the discrepancies between reports from children and parents. The study focused on several possible factors which may contribute to the discrepancies in the reports obtained from children and parents in the community. These included:

1. The age of the child.
2. The nature of the child's symptomatology.
3. The presence of parental psychopathology.

4. The psychological adjustment of the child's family.

A secondary aim of the study was to compare the pattern of discrepancies between parent and child reports describing children living in the community, with the pattern of discrepancies between reports describing children referred to mental health clinics.

The unique features of the study are the combination of comprehensive data describing the participating children, obtained from four different informants in most instances, the inclusion of children in an age range where the reliability of the children's reports is acceptable (Edelbrock et al., 1985), the inclusion of representative samples of children from the general community and from mental health clinics, and finally, the use of similar measures to obtain reports about the children from four different informants.

CHAPTER 2. STUDY DESIGN AND METHODOLOGY

2.0 INTRODUCTION

This chapter describes the method of sample selection, the method of data collection, and the measures used in the study. In addition, an outline is provided of the statistical methods used in the study.

2.1 SAMPLE SELECTION

Subjects in the community sample consisted of a randomly selected group of 696 children in two specifically defined age groups living in the community in Adelaide, South Australia (Table 2.1). The strategy of studying children in the general population was adopted in order to avoid the misleading effects of confounding variables which can arise when only clinic samples are studied. As Henderson et al. (1981) have emphasized, "there are often substantial differences in the social and intrapersonal attributes of those who are in contact with doctors or clinics and those who have the same symptoms or disorders but have not consulted medical agencies." (p. 85). A separate study of a clinic sample of eighty-two children, aged 10-16 years, who had been referred to a mental health clinic in Adelaide (Table 2.1) was undertaken in order to determine whether results from the community children were different from those which would be found with clinic-referred children. Results from this study are reported in Chapter 7.

The sample of children from the general population included children aged 10-11 years and 14-15 years. These age groups were chosen for several reasons. First, children in these age ranges typically are still living at home and assessment of emotional and behavioural problems relies on reports from parents, children and teachers. Second, the approach of including children from two age groups, which roughly approximate to pre-puberty and post-puberty, allowed comparisons to be made between reports describing children at different stages of their development. At the same time, limiting the study population to children of specific ages ensured that the size of the total sample was kept to manageable proportions. Finally, the lower age limit of 10 years was chosen because there is evidence (Edelbrock et al., 1985) that information obtained from younger children is less reliable than that obtained from children over the age of 10 years.

The initial sample of 696 children included 358 who were aged 10-11 years and 338 aged 14-15 years. The two cohorts were selected from the total population of children of the appropriate age enroled in schools in metropolitan Adelaide. Although accurate figures are not available, it is believed that less than 1% of children in these age ranges are not enroled in school (M. Caust, Manager, Research and Statistics, South Australian Education Department, personal communication, February 1st 1988).

A stratified sampling procedure was used, as a separate aim of the project was to compare the prevalence of emotional and behavioural problems in schools of differing socio-economic class. All schools (including private and public schools) in the Adelaide metropolitan

region were initially ranked on the basis of the number of children in the school receiving benefits from the Government Assistance Scheme (The South Australian Education Department uses this criterion to identify the socio-economic class of schools). The ranked list of schools was then divided into three strata, each containing an equal number of schools. Schools were randomly selected from within each stratum and children randomly selected from within each identified school. Three criteria were used to determine both the number of schools and children selected. First, the maximum proportion of children selected from each school was set at 1 in 5. This was done to ensure that no single teacher was asked to complete more than five teacher checklists (on average, classes in the participating schools contained 25 children). This number of checklists was selected after discussion with teachers who suggested that five checklists was the maximum which an individual teacher could reasonably be expected to complete for the study. Second, the sample size was chosen on the basis that it would be large enough to detect a difference of half a standard deviation between the schools of different socio-economic class using the Child Behavior Checklist (Achenbach & Edelbrock, 1983). This difference was chosen because, if present, it represented a clinically significant difference between the number of problems in children in the three school groups. Finally, the total sample size was chosen with the aim of meeting the criterion (Boyle et al., 1987) of obtaining 95% confidence intervals whose width is approximately 8% of the anticipated estimates.

A total of eleven secondary schools and twenty primary schools were selected to participate in the study. All the schools that were approached agreed to participate and actively co-operated throughout the study. The vast majority of the primary school children were enroled in

Grades 5-6 and the secondary school children in Grades 9-10. There was only one family from which two children were selected (by chance) and one of these children was randomly chosen to be dropped from the study.

2.2 DATA COLLECTION

A mail survey was the principal technique used for the collection of data from participating families in the community sample. In addition, 200 of these families were randomly selected for home visits (Table 2.1). Data from the clinic families was all collected during visits to the families homes.

2.21 Rationale for using a Mail Survey

Mail survey techniques provide an economical and efficient means of surveying large community samples and Dillman (1978) has shown that very high response rates can be obtained from randomly selected community samples using this methodology. In addition, as highlighted by Horwath (1987), mail surveys provide "far less opportunity for the investigator to bias the respondent, and respondents are more clearly volunteers" (p. 71). However, it should be noted that there is still the possibility of a "self-selection" bias as potential participants clearly have the opportunity to decide whether or not they will participate in a mail survey.

One limitation of mail surveys is the possibility that answers provided by respondents have either been discussed with other family members or are influenced by their presence. This was of particular concern in this

study as it aimed to collect independent reports about children's behaviour from different members of the same family. Two different approaches were used to try to ensure that family members independently completed their questionnaires. First, the questionnaires used were highly structured. Previous studies have shown that there are few differences in the pattern of responses and the quality of data obtained between face to face, telephone and self-administered modes of questionnaire completion, provided that a highly structured questionnaire is used (Quine, 1985). Second, great emphasis was placed in the letter sent to families, and in the introduction to the questionnaires on the importance of family members providing independent reports.

Two hundred children in two-parent families who took part in the mail survey were randomly selected for home visits (Table 2.1). There were two reasons for conducting these visits. First, the visits were used to provide a check on the independence of reports obtained in the mail survey from different members of the same family. This was done by comparing results obtained in the mail survey with results obtained on the same questionnaires administered independently to each family member by a research assistant during a home visit (see Chapter 8). Second, it was considered important that interview-based assessments be included in the study as this is the primary assessment method used by clinicians when evaluating children with emotional and behavioural problems. However, a major problem with using parent and child interviews to assess large numbers of children is that interviews are very time consuming and labour intensive. In order to achieve the dual aims of including interview-based assessments, while at the same time keeping the overall size of the study to manageable proportions, 100 children were randomly selected from amongst the 200 children chosen for a home visit, for

assessment by means of the Diagnostic Interview Schedule for Children and Parents (DISC) (Costello et al., 1984) (Table 2.1). All the children and parents in the community families who participated in the home visits were visited after they had completed the mail survey.

Results obtained from the parents and children in the community sample, who participated in the home visits were compared with the results obtained from the parents and children in the clinic sample (Table 2.1). This comparison was undertaken in order to determine whether the level of agreement between parents and children in families where the child had been referred to a mental health clinic differed from that of parents and children living in the general community. None of the participating children in the community families selected for a home visit had been referred to a mental health clinic. Full details of the clinic sample are provided in Chapter 7.

2.22 Mail Survey Protocol

The mail survey protocol was based on that designed by Dillman (1978) and Dillman's recommendations for questionnaire design and follow-up contacts were followed during the mail survey component of the study. The schedule employed during the survey is shown in Figure 2.1.

The initial mailing included a cover letter, separate questionnaires for mothers, fathers and children, and a reply paid envelope in which to return the questionnaires. The cover letter described the aims of the study, assured confidentiality and provided a contact number for queries. In addition, letters emphasised the importance of each family member independently completing their own questionnaire. Each letter was

individually signed in accordance with Dillman's instructions. One week after the initial mailing a reminder card was sent to all families. The card thanked families who had already returned their questionnaires and encouraged those who had not done so to return their questionnaires. Two weeks later a telephone call was made to all families who had not returned their questionnaires. Two attempts were made to establish telephone contact with each family, one call made within normal working hours and one in the evening or on the weekend. Telephone contact was attempted with 283 families and contact made with 199 families. A further reminder letter containing a new set of questionnaires was sent one week after the telephone call and a final reminder letter two weeks later. These reminder letters emphasized the importance of each individual family's participation for the success of the study.

The daily return rate of questionnaires is shown in Figure 2.1. It can be seen that approximately 50% of questionnaires were returned within two weeks of being mailed out. The telephone call reminder played an important role in answering questions and encouraging the return of questionnaires. Subsequent reminders by letter appeared to play a lesser role but did result in the return of a number of questionnaires.

As promised, all participating families and the headmasters of participating schools were sent a summary of results from the study in July 1988.

2.3 MEASURES

The measures used in the study are shown in Table 2.2.

2.31 Child Behaviour Checklists

All participating children were assessed by means of the Child Behavior Checklist (CBCL) (Achenbach & Edelbrock, 1983) completed by parents, the Youth Self Report (YSR) (Achenbach & Edelbrock, 1987) completed by children, and the Teacher Report Form (Achenbach & Edelbrock, 1986) completed by teachers. These checklists were chosen because they are designed to be self-administered, provide comparable information from parents, children and teachers, and there is extensive data supporting their reliability and validity (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986; Achenbach & Edelbrock, 1987). Prior to the study commencing, advice about the use of the checklists was obtained directly from Professor Achenbach.

A major advantage of the CBCL, YSR and TRF is that they employ the same answering format and they employ very similar items to describe children's behaviour. As mentioned, previous research investigating discrepancies between reports obtained from different informants has often used different measures to obtain reports from the different informants. This makes it difficult to know whether differences identified between reports are due to the use of different measures, i.e. method variance, or are due to differences in the characteristics of the informant, i.e. informant variance.

The CBCL is designed to collect information from parents in a standardized way about children aged 4-16 years. The checklist consists of two parts. The first part of the checklist is designed to obtain information about the social competencies of children. Parents are asked to describe their child's participation and skills in activities and sports, and their degree of participation in social organizations. In addition, information is obtained about children's contacts and behaviour with other children, and their competence at school. The second part of the checklist consists of 118 items describing a wide range of behaviour problems. Parents are asked to report on their child's behaviour "now or within the past 6 months" using a three point scale (0 = Not true, 1 = Somewhat or sometimes true, 2 = Very true or often true). Principal components analyses of results obtained from 2,300 children referred to 42 mental health services in North America identified either 8 or 9 "narrow band" factors for children in three age groups (4-5 years, 6-11 years, and 12-16 years). Second-order principal components analyses of these factor-based narrow band behaviour problem scales identified two second-order factors which Achenbach and Edelbrock labelled "Internalizing" and "Externalizing". These latter factors reflect a "distinction between fearful, inhibited, overcontrolled behavior and aggressive, antisocial, undercontrolled behavior." (Achenbach & Edelbrock, 1983 p. 31). A Total Problem Score comprising all the problem items on the checklist is also scored on the CBCL. The CBCL has been widely used in American, Dutch and Australian studies for studies of both community and clinic-referred children (Verhulst, 1985; Hensley, 1988; Bird et al., 1988).

The YSR and TRF are very similar to the CBCL. The YSR contains the same number of items as in the CBCL, however, these include 16 socially desirable items which replace items in the CBCL which Achenbach felt were inappropriate to ask adolescents. The socially desirable items are completed by adolescents but are not included in the scoring for the YSR. Eighty-nine items are the same on the CBCL, YSR and TRF. The YSR is designed to collect information from children aged 11-18 years and the TRF is designed to collect information from teachers about children aged 6-16 years. Like the CBCL, the YSR and TRF consist of two parts, the first rating children's social competencies and the second, children's behaviour in a range of areas. Once again, principal components analyses were used to identify "narrow band" factor based behaviour scales on the two checklists and second-order analyses showed that these factors formed two broad-band groupings which were labelled "Internalizing" and "Externalizing" (Achenbach & Edelbrock, 1986, Achenbach & Edelbrock, 1987). Considerable information describing the validity and reliability of the YSR and TRF is available (Achenbach & Edelbrock, 1986, Achenbach & Edelbrock, 1987)

2.32 General Health Questionnaire

The psychological adjustment of parents was assessed by means of the General Health Questionnaire (GHQ) (Goldberg, 1978). Two different versions of the GHQ were used in the study, the twelve item version (GHQ-12) and the thirty item version (GHQ-30). All parents in the community families completed the GHQ-12 which was distributed as part of the mail survey. Parents who participated in the structured interviews completed the GHQ-30.

The GHQ is a self-administered questionnaire which is designed to identify individuals with non-psychotic psychological impairment. Several different versions of the GHQ are available consisting of 12, 20, 28, 30, and 60 items. In each version a four point answering system is provided consisting of "Better than usual, Same as usual, Worse than usual, Much worse than usual". Goldberg (1978) states that the GHQ "concerns itself with two major classes of phenomena: inability to continue to carry out one's normal "healthy" functions, and the appearance of new phenomena of a distressing nature" (p. 5). It is "specifically concerned with the hinterland between psychological sickness and psychological health" (p. 6). A particular advantage of the GHQ is that it has been widely used in Australian populations (Henderson et al., 1981; Tennant, 1977) and Australian norms are available. There are two principal shortcomings of the GHQ. First, individuals with long standing problems may correctly not endorse items which they have experienced "no more than usual" thereby resulting in chronic problems being missed. Second, mentally healthy individuals with a physical disorder, or women who are pregnant, may endorse some items thereby generating false positive scores.

The decision to use the 12-item version of the GHQ-12 in the mail survey was based primarily on the need for brevity. Although there is some decline in the sensitivity, specificity, and misclassification rates of the 12-item version when compared to longer versions of the GHQ (Tennant, 1977), it still achieves a high standard in these areas. In the structured interviews the GHQ-30 version was chosen because of its superiority over other similar instruments (Henderson et al., 1981, p. 91). In this thesis, only results from analyses using the GHQ-12 are reported.

2.33 Family Assessment Device

Family adjustment was assessed by means of separate reports obtained from the mother, father, and the 14-15 year old children, using the General Functioning Scale of the Family Assessment Device (FAD-GF). Pilot testing had shown that some 10-11 year old children were not able to complete independently the FAD-GF and it was not included in the booklets for these children in the mail survey.

The FAD (Epstein, Baldwin & Bishop 1983; Miller, Epstein, Bishop & Keitner 1985; Sawyer, Sarris, Baghurst, Cross, & Kalucy, 1988) is a 60-item self-report questionnaire developed to measure family health/pathology which can be completed independently by family members over the age of twelve years. It is based on the McMaster Model of Family Functioning (Epstein et al., 1983) and focuses on six specific dimensions of family functioning labelled: "Problem Solving", "Communication", "Roles", "Affective Responsiveness", "Affective Involvement", and "Behavior Control". In addition, there is a "General Functioning" scale (FAD-GF) which provides a measure of the overall health/pathology of the family.

The FAD-GF was derived from item analyses conducted during testing of the psychometric properties of the FAD. The FAD had been designed with the intention that each item would apply to a single dimension of the McMaster Model of Family Functioning (Epstein et al., 1983; Miller et al., 1985). The initial version thus contained six sets of items, but the scores from the six sets were found to be highly intercorrelated. In order to overcome this problem, the most highly intercorrelated items

were selected out to form the FAD-GF. When the shared variance between the FAD-GF and the other subscales was removed statistically, intercorrelations between the other subscales were found to be acceptable. Scores on the FAD-GF range from 1-4 with higher scores indicating less healthy functioning.

Byles, Byrne, Boyle, and Offord (1988) have reported on the value of the FAD-GF as a measure of family functioning in survey research. In their study, Byles et al. (1988) evaluated both the validity and the reliability of the FAD-GF during its use in a large scale epidemiological study of children and families. Concurrent validity of the FAD-GF was established by showing that scores on the FAD-GF were significantly associated with other variables (for example, marital separation and alcoholism) commonly considered to be associated with family dysfunction. The internal consistency of the FAD-GF " ... was found to be .86 (Cronbach's alpha) and the split-half coefficient (Gutman) was .83" (Byles et al., 1988, p. 102).

2.34 Diagnostic Interview Schedule for Children and Parents

The Diagnostic Interview Schedule for children (DISC) is a highly structured interview format which was developed by Costello et al. (1984) for use in community studies of psychopathology in children aged 6 to 18 years.

The DISC consists of parallel interviews for children (DISC) and parents (DISCP) which can be given by trained lay interviewers or clinically trained personnel. It incorporates features of previous clinical and epidemiological interviews and covers most childhood mental disorders.

The time span for questioning about symptomatology generally covers the year prior to assessment although, on occasion, other specific time prompts are used. The DISC employs a no/sometimes/yes answering format that corresponds to a 0,1,2 coding pattern. Considerable data supporting the validity and reliability of the DISC and DISC-P are available (Costello et al., 1984). For the purpose of this study, only items comprising the depression, anxiety, oppositional disorder and conduct disorder scales were included in the interviews. These areas were chosen because they provided a balance of potentially observable symptoms, and symptoms of a more private nature which might be less obvious to an outside observer such as a parent. A Total Symptom Score was obtained by summing the scores from all four scales. As well, a Behaviour/Conduct Score was obtained by summing the scores from the conduct and oppositional scales, and an Affective/Neurotic Score was obtained by summing the scores from the depression and anxiety scales. When the study commenced, changes were impending to the specific criteria used to define many childhood psychiatric disorders. As a result it was decided not to use the diagnostic algorithms available with the DISC as it was anticipated that they would be superseded by others before the study was finished.

Prior to the commencement of the study, Professor Costello was consulted about the use of the DISC. For the first year of the study two research assistants were employed on a half-time basis to conduct the structured interviews. Both research assistants were trained in the use of the DISC before commencing to use it. In addition, a study of their interrater reliability was undertaken using both 10-12 year old and 14-15 year old children and their parents. The results of this study are reported in the Appendix and it can be seen that high levels of interrater

reliability were attained. As a result of experience gained in the study of interrater reliability, the questions from the Oppositional Symptom Score were added to the structured interview as it was apparent that few of the questions in the Conduct Disorder area were being endorsed by either children or parents. After one year, one of the research assistants left the study and the other assistant then continued on a full-time basis and conducted all remaining interviews.

2.35 Pilot Testing of the Questionnaires

The questionnaires used by children in the mail survey were all pilot tested with children of the appropriate age from two primary schools and a secondary school. Twenty 10-11 year olds participated in the pilot testing and a similar number of 14-15 year olds.

The pilot testing particularly focused on the 10-11 year old children as it was hoped that, despite their young age, they would be able to complete independently the Youth Self Report and the General Functioning Scale of the Family Assessment Device. In fact, testing showed that some of the younger children were not able to complete independently the General Functioning Scale of the Family Assessment Device, nor were they able to complete the competence items from the Youth Self Report. In each case the children had difficulty using the answering formats without help from an adult. All the children were able to understand and complete the items describing behaviour problems on the Youth Self Report. In view of this, the General Functioning Scale was not included in the mail survey questionnaire for 10-11 year old children, and only the behaviour problem items were included from the Youth Self Report.

2.36 Questionnaire Design for the Mail Survey

The questionnaires used in the mail survey (Table 2.2) were all bound into attractively covered booklets. Different colours were used for booklets completed by mothers, fathers and children. A short introduction was included at the beginning of each booklet, including a clear statement indicating which family member was to complete the booklet. A question asking the respondent to identify their sex was included amongst questions in the booklets designed for parents. This made it possible to identify booklets which had been inadvertently answered by the wrong parent.

2.4 STATISTICAL METHODS

The specific statistical techniques used for the analyses of results are described in the text at the point where they are employed.

In some instances, multiple comparisons were made between informants describing children in the study. For example, when comparisons involved the Externalizing, Internalizing, and Total Problem scores on the parent and child behaviour checklists for the four groups of children (male and female children, aged 10-12 and 14-15 years old), there were a total of 12 comparisons in each set of analyses. Feild and Armenakis (1974) have pointed out that in such circumstances, the probability that the null hypothesis will be rejected, i.e. there are no differences between the groups, increases as the number of independent statistical tests increases.

Feild and Armenakis (1974) identified three approaches which have been used with multiple comparisons:

1. Proceeding with the multiple-significance tests without taking into account the earlier mentioned cautions. If this is done, "the investigator does not operate at his *a priori* level of significance" (p. 430).
2. Proceeding with the multiple-significance tests but to regard as significant only those differences which are significant at a level lower than the conventional 0.05 level. A disadvantage of this option is that the probability of accepting the null hypothesis when it is false increases (Type II error), "hence there is a loss of power" (p. 430).
3. Initially undertaking an over-all test on the data and if this is significant, proceeding with multiple-comparison tests in order to isolate specific group differences. Consistent with most standard statistical texts, this is identified as the "most advantageous" (p. 430) alternative.

For the purpose of analyzing results, the third alternative was generally employed in this study. However, where multiple comparisons were undertaken, the number of $p<0.05$ findings expected by chance using a $p<0.05$ protection level is identified in the text where the comparisons are reported. For example, in a set of twelve comparisons, two to three significant findings could be expected by chance using this level of protection (Feild and Armenakis, 1974). In light of this, if no more than three of a set of twelve comparisons yielded significant ($p<0.05$) results, this could be considered a chance finding.

Finally, it should be noted that when results are presented in tables, percentages may not add to exactly 100 per cent due to rounding. In general, the number of subjects missing are identified in the tables except where the number missing was very small.

CHAPTER 3. DEMOGRAPHIC AND PSYCHOLOGICAL CHARACTERISTICS OF THE STUDY POPULATION

3.0 INTRODUCTION

This chapter is divided into three sections. The response rates obtained in the study and possible selection biases are reported in the first section. The demographic characteristics of the study population are described in the second section.

In the final section, the psychological adjustment of the study population is described. In this section, scores were adjusted to take into account the stratified sampling procedure used in the sample selection. This adjustment was done so that results obtained in the study could be compared with those of previous studies in Australia and North America. It should be noted, however, that differences between the adjusted and unadjusted scores were minimal (less than 3% of the adjusted score).

3.1 THE STUDY POPULATION

The number of informants in each of the four groups (10-11 year old male children, 10-11 year old female children, 14-15 year old male children, 14-15 year old female children) are identified in Table 3.1, and the response rates for different informants are shown in Table 3.2. Overall, a response from a child and at least one parent was achieved with 76% of

the 10-11 year olds and 71% of the 14-15 year olds. It can be seen in Table 3.2 that the best response rates were obtained from mothers and children. The poorest response rates occurred with teachers and fathers of the older children. In general, better response rates were obtained from informants in the female than in the male groups. The response rates identified for fathers in Table 3.2 take into account the number of families in which it was known from the mother's demographic questionnaire that there was no father in the family. Even so, it is likely that the response rates identified for fathers are an underestimate of the true picture as some children for whom no response was received from any family member would have been living in single mother families.

Four different approaches were taken to identify possible selection biases in the study population. Firstly, the content of all telephone contacts with parents during the study was recorded in order to try to identify possible biases in the study population arising because of the non-participation of some families. Table 3.3 summarizes the content of these contacts.

There was no single predominant reason given by parents for non-participation in the study. Most parents said that they simply did not wish to participate. Stated reasons for non-participation included a feeling that the study represented an invasion of privacy (N=5), the family was too busy to complete the questionnaires (N=4), the parents did not speak English well enough to answer the questionnaires (N=4), and that the content of the questionnaires was too negative (N=3). One home visit was undertaken after the parents, whose knowledge of English was

poor, asked that someone come to their home to help with answering the questionnaires.

In six of the fourteen families in which a parent telephoned specifically to say that they did not wish to participate, it appeared that the child may have been at risk for emotional and behavioural problems. For example, one parent specifically said that she had problems with her child, another said that her marriage had recently broken up, and another did not wish to participate in the study but wanted psychiatric help for her child. Although this suggested the possibility that children with emotional and behavioural problems may be under-represented in the community sample, the risk factors did not appear to be present in families who indicated at the time of the later telephone reminder call that they did not wish to participate in the study.

Secondly, the response rates for the informants in each socio-economic school group were identified and are shown in Table 3.4. It can be seen that for all informants, the best response rates were obtained from the highest socio-economic class school group and the worst response rates were from the lowest socio-economic class school group. This suggests that the study population may be under-representative of children from lower socio-economic class families.

Thirdly, in order to try to determine whether the teacher reports (which showed the lowest response rates) described a biased group of children in the study population, the mother, father, and child total problem scores (CBCL and YSR) describing children for whom a teacher report was available, were compared with the mother, father, and child scores for the children where the teacher reports were missing. The only

significant ($p<0.05$) differences between the two groups occurred with the 10-11 year old male children. For these children, the mothers ($t=2.2$, $df=126$, $p<0.03$) and the children ($t=3.2$, $df=129$, $p<0.002$) reported that the children for whom teacher reports were available had significantly fewer problems than the children for whom teacher reports were missing. This suggests that teacher reports in the group containing the 10-11 year old male children may be under-representative of children with behaviour problems.

Finally, where data was available, the characteristics of the study population were compared to those of the population of the Adelaide Statistical Division at the nearest population census (1986). These comparisons are shown in Section 3.2 below. The possible influence of particular sample biases on the results of the study are discussed in the text where the results are presented.

3.2 DEMOGRAPHIC CHARACTERISTICS OF THE STUDY POPULATION

3.21 Age and Sex

The sample consisted of children aged 10-11 years and 14-15 years on July 1st 1987 (Table 3.5). The male to female ratio in the sample (Table 3.1) closely resembled that of the Adelaide Statistical Division. For the younger group the ratio was 0.90 and the older group it was 1.03. The equivalent ratios for the Adelaide Statistical Division were 1.07 and 1.05. (Australian Bureau of Statistics, 1986).

3.22 Domestic Situation

In each of the four age and sex groups, approximately 90% of children lived in families where there were two parents (Table 3.6). The majority of single parents were mothers with only a very small percentage of single parents being fathers. There was no significant difference between the four sex and age groups with regard to the proportion of two parent and single parent families in each group (Likelihood Ratio $X^2=5.3$, df=6, p=0.5).

The percentage of single fathers in the study is very similar to that reported for the Adelaide Statistical Division (2%), however, the percentage of single mothers, especially of the female children is less than that reported for the general population (14.5%). A relatively poor response rate from single mothers appears to be the principal reason for this difference. In the sample of children originally selected for the study, a mother was the only parent identified on the school lists for 13.4% of the children, a prevalence which was consistent with the prevalence of single mothers in the Adelaide Statistical Division. However, the response rate from this group of mothers was only 53.8%, considerably lower than the response rate for the total sample (74.1%). The records kept of contacts with parents suggested that single mothers felt they had insufficient time available to answer questionnaires.

3.23 Occupational Class of Parents

The distribution of occupational class for both mothers and fathers in the study groups is reported using Daniel's (Daniel, 1983) Australian

occupational status scale in Tables 3.7 - 3.8. This scale was developed with the intention that it reflect the prestige of a wide range of occupations in Australia. As a guide, 1 to 2.9 equates to occupations such as medical specialist, barrister, and bank manager, 3 to 4.9 includes occupations such as computer programmer and electrician, and 5.0 to 6.9 includes occupations such as typist, junior clerk and barman. One limitation of the scale is that it does not include a rating for "Housewife" or "Homeduties". As a large percentage of mothers in this study identified "Homeduties" as their occupation, it is separately identified in Table 3.7. There was no significant difference between the four age and sex groups in the distribution of occupations reported by mothers (X^2=9.3, df=9, p=0.4) or fathers (X^2=12.7, df=9, p=0.2). Due to the incompatibility of the classification system used by the Australian Bureau of Statistics and Daniel's occupational status scale it was not possible to make meaningful comparisons between the distribution of occupations in this sample and distributions in the general population.

3.24 Educational Attainment

The level of education completed by mothers and fathers is shown in Tables 3.9 - 3.10. There was little difference in the education of mothers in the four age and sex groups (X^2=8.9, df=12, p=0.7). For fathers, the significance level of the Chi-square statistic testing for equality of proportions in the four groups was p=0.06 (X^2=20.2, df=12, p=0.06). This reflected some differences in the educational achievements of fathers in the different age and sex groups. For example, a higher proportion of the fathers of the male children had not progressed beyond primary school education. As well, a higher proportion of fathers in

the group containing the older female children had obtained tertiary qualifications compared to fathers in the other groups.

In general, the educational experience of the parents did not differ greatly from that of the total population of the same aged adults (35-44 years) living in the Adelaide Statistical Division (Australian Bureau of Statistics, 1986). The one consistent difference was that a higher proportion of parents in the study had tertiary qualifications. A possible explanation for this finding is that better educated parents may have been more likely to respond to the mail survey, which required completion of a detailed questionnaire, than less educated parents. This explanation is consistent with the observation that better response rates were obtained from parents whose children were attending the higher socio-economic class schools.

3.25 Country of Birth

The distribution of parents by country of birth is shown in Tables 3.11 - 3.12. The distribution in the study is very similar to that reported for adults aged 25 - 54 years in the Adelaide Statistical Division. The largest difference occurred with the fathers in the group containing the 10-11 year old females. In this group, a smaller proportion of fathers than in the other groups reported their place of birth as Australia or New Zealand (in order to enable comparison with the population in the Adelaide Statistical Division, Australia and New Zealand are combined in Tables 3.11 - 3.12) and a larger proportion reported their country of birth as UK/Ireland.

The results of statistical analyses suggest that the proportion of fathers (X^2=18.7, df=12, p=0.1) born in different countries varied somewhat in the four age and sex groups. Less variation was evident between mothers in the four groups (X^2=15.5, df=12, p=0.2). As shown in Table 3.13, for parents born overseas, the average length of time the parents had lived in Australia was 20 years. This mitigates against country of birth having a strong effect on the responses from parents in the study groups. Once again, although there were some variations between groups, the overall distribution for parents in the study was similar to the total population of 35-44 year olds in the Adelaide Statistical Division (Australian Bureau of Statistics, 1986).

3.3 PSYCHOLOGICAL ADJUSTMENT OF CHILDREN, PARENTS AND FAMILIES

3.31 Psychological Adjustment and Competencies of the Children

A) Parent Report: The mean scores obtained from mothers and fathers in the study (Table 3.14) were higher than those reported (largely by mothers) by Achenbach and Edelbrock (Achenbach & Edelbrock, 1983) for North American children, but lower than those obtained by Hensley (Hensley, 1988) in her study of children in Sydney. One possible explanation for the different results in this study is that it is less representative than the Sydney or North American studies and some sample selection may have occurred. However, the mean scores reported in this study are generally comparable with the other studies which have used the CBCL (Achenbach & Edelbrock 1983; Hensley, 1988; Verhulst, 1985; Bird et al., 1988) and it is reassuring that they lie between those found in the studies of Achenbach and Edelbrock (1983), and Hensley (1988).

Results from the Social Competence Scales of the Child Behavior Checklist in this study (Table 3.15) are similar to those of Achenbach and Edelbrock (1983), and Hensley (1988). Interestingly there is much greater consistency in the results from the three studies in this area than in ratings of behaviour problems.

B) Child Self-Report: Mean scores from the Youth Self Report are shown in Table 3.16. No other comparable Australian scores are available, however, mean scores on the Youth Self Report reported by Achenbach and Edelbrock (1987) are shown in the table. In contrast to scores on the behaviour scales obtained from parents, the scores obtained from the children in Adelaide were generally somewhat lower than those reported by the children in North America. This was particularly evident for the female children. Previous studies in different countries have reported varying YSR scores. For example, lower YSR scores were also found with community samples of adolescents in Puerto Rico and Holland, but similar scores to the U.S.A. were found in West Germany (T. M. Achenbach, personal communication, October, 1989).

Mean scores from the Social Competence Scales on the Youth Self-Report completed by the 14-15 year old children are shown in Table 3.17. Once again, results from the Social Competence Scales were very similar to those reported by Achenbach and Edelbrock (1987). This suggests that children in Adelaide viewed their social competencies as being similar to those of their counterparts in North America but they appeared to perceive themselves as having fewer emotional or behavioural problems.

C) Teacher Report: Mean Scores from the Teacher's Report Form are shown in Table 3.18. The mean scores were less than those reported by Achenbach and Edelbrock (1986), except in the case of the group containing the 10-11 year old male children. As noted earlier, the response rate from teachers was relatively low in this study and it is possible that there may have been some biased sample selection in this group with omission of teacher reports from children with a greater number of problems.

Mean scores from the adaptive functioning scales of the TRF are shown in Table 3.19. It can be seen that, in contrast to scores from the behaviour scales, the differences between competence scores obtained from children in this study and those of the North American children were small. Except in the case of the group containing the 14-15 year old female children, teachers in this study scored the children as more competent than their North American counterparts.

D) Prevalences: Achenbach and Edelbrock (1983) identified cutoff scores which marked the "limits of the 'normal range' for each sex/age group" (p. 63). These scores equated to the 90th percentile of the Total Behavior Problem score (Achenbach & Edelbrock, 1983, p. 62). In this thesis, the term "cases" will be used to refer to those children whose Total Behavior Problem score exceeded the cutoff applying to their sex/age group.

Higher mean scores on the CBCL which were obtained in a recent epidemiological study of Australian children (Hensley, 1988) suggest that somewhat higher cutoff scores may be appropriate for Australian children.

To date, although there is little information about scores which apply to clinically referred children in Australia, Hensley (1988) has suggested that the raw scores which equate to the 90th percentiles in her study could be used as tentative cutoff scores for Australian children living in urban settings. Hensley (1988) noted that if her recommendation was followed, it "would considerably reduce the number of children regarded as being in the clinical range" (p. 379). In light of these differences, prevalences in this study were calculated separately using the cutoff scores recommended by Achenbach and Edelbrock (1983, 1986, 1987), and the cutoff scores suggested by Hensley (1988) for use with Australian populations (Table 3.20). The prevalences were all adjusted to take into account the sampling procedure.

It can be seen that there were considerable differences in the prevalences identified using the different cutoff scores available for the CBCL. It is interesting to note that, although the mean scores in this study were considerably lower than those identified by Hensley (1988), the prevalence of cases identified (from mothers' reports) using the cutoff scores recommended by Hensley (1988) were only slightly less than 10%. This suggests that the number of children reported as having extreme scores in this study were similar to the number in the study by Hensley (1988). In contrast, when the cutoff scores recommended by Achenbach and Edelbrock (1983) were employed, much higher prevalences were identified.

Finally, it is important to note that there were large differences in the prevalence of cases identified by parents, children, and teachers. The differences in the number of cases identified by different informants is described in greater detail in Chapter 5.

3.32 Psychological Adjustment of the Parents

The four-point response scale on the GHQ can be scored in two ways. Scores can be treated as a multiple-response or "Likert scale" or alternatively scores can be treated as a bimodal response scale (Goldberg, 1978). Results from the GHQ-12 obtained using both these scoring methods are reported in Tables 3.21 and 3.22. In addition, the recommended cutoff scores were used to identify the prevalence of mothers and fathers ("cases") who scored in the clinical range.

Prevalences found in this study were similar to those found in previous studies of Australia populations which have used the GHQ-12 (Burvill & Knuiman, 1983). In three of the four study groups, the prevalence of cases amongst the fathers was lower than that for the mothers. This pattern of lower prevalences for males has been reported in previous studies in Australia. For example, Shiraer and Armstrong (1975, p. 239) reported a prevalence of 26.7% cases amongst men and 27.7% amongst women in Gosford/Wyong, and 18.4% for men and 26.3% for women in Illawarra, New South Wales.

3.33 Family Adjustment

The psychological adjustment of families was assessed by means of the General Functioning Scale (FAD-GF) of the Family Assessment Device. Reports were obtained from both parents and the older children (Table 3.23).

The FAD-GF scores obtained from the participants in this study were very similar to those which were obtained from community families in an earlier study which had evaluated the validity of the FAD as a measure of family adjustment. (Sawyer et al., 1988). The earlier study compared FAD scores obtained from family members in families living in the community, with scores obtained from family members in families where a child had been referred to a mental health clinic. In the earlier study, subjects in the community families rated their families as significantly better adjusted than did members of the clinic families. Taken together, the results from the earlier study and this study suggest that the psychological adjustment of the community families in this study is typical of that of families living in the community in Adelaide. In addition, the combined results suggest that the community families in this study were functioning better than families in which a child had been referred to a mental health clinic.

The only comparable study which has used the FAD-GF is the study by Byles et al. (1988). Although it is unclear which parent completed the FAD-GF in the study by Byles et al. (1988), the mean score reported for the FAD-GF (Mean Score=1.75, SE=$\pm$0.01) in that study is comparable with the score obtained from parents in this study. It is interesting to note that the children in this study and in the earlier study rated their families as less well adjusted than did their parents. In both studies, the mean scores from the fathers tended to occupy an intermediate position between the mean scores reported by the mothers and children.

CHAPTER 4. COMPARISON OF THE REPORTS DESCRIBING CHILDHOOD EMOTIONAL AND BEHAVIOURAL PROBLEMS OBTAINED FROM MOTHERS, FATHERS, CHILDREN, AND TEACHERS

4.0 INTRODUCTION

Clinical assessment of childhood emotional and behavioural problems relies heavily on reports from mothers, fathers, children, and teachers. A limitation of earlier studies is that they have generally included reports from only a single pair of informants, for example, mothers and children (Herjanic & Reich, 1982; Edelbrock et al., 1986), or mothers and fathers (Ferguson et al., 1974; Jacob et al., 1982). As a result, there is a lack of information about the level of agreement between the four informants traditionally employed in the assessment of childhood problems.

A further limitation of earlier studies is that different measures were often used by different informants to describe children's behaviour (Anderson et al., 1987). This makes it difficult to know whether differences identified between informant reports were due to differences in the measures used, or differences in the characteristics of the informants. Finally, previous studies have varied greatly in the age of participating children. This is of particular concern in studies which have included relatively young children because Edelbrock et al., (1985) have pointed out that the reliability of reports from children under the age of 10 years is relatively low.

The results in this chapter compare reports describing childhood emotional and behavioural problems obtained from parents, teachers, and children living in the general community. The children were aged 10-11 years and 14-15 years at the time of the study and similar measures were used to obtain reports from all the informants. The results focus both on the number of problems reported by the different informants and on the level of consistency between their reports.

4.1 PROCEDURE

The Child Behavior Checklist (CBCL), Youth Self Report (YSR) and Teacher Report Form (TRF) were used to obtain reports about the children's behaviour from parents, children, and teachers.

Despite their similarity, there remain a few important differences between the three checklists which must not be overlooked. First, there are differences in the item content of the checklists (see Appendix). Sixteen items in the CBCL which "were deemed inappropriate to ask adolescents, mostly because they are characteristic of younger ages." (Achenbach & Edelbrock, 1987, p. 7) were not included in the YSR. These items were replaced with socially desirable items which are not scored on the YSR profile. A further thirteen items on the CBCL are not present on the TRF and are replaced with items describing school related behaviour (Achenbach & Edelbrock, 1986). Second, there are different narrow-band factors on the different checklists and within the same checklist the Externalizing and Internalizing Factors are comprised of different items for children of different age and sex. Finally, there is a difference in

the time period over which teachers are asked to base their ratings of children's behaviour on the TRF. Teachers base their ratings of children's behaviour on observations over a 2 month period while parent and child ratings are based on observations over 6 months.

The differences between the item content of factors on the three checklists make it difficult to compare the total number of problems reported by different informants describing children of different age and sex. Although T Scores are available which provide "a common metric across scales differing in raw scores" (Achenbach & Edelbrock, 1983, p. 26), a single T Score may equate to more than one raw score on a single checklist and the same T score often equates to different raw scores on different checklists.

For the purpose of comparisons reported in this chapter, two different approaches were used to ensure that the Total Behaviour Problem scores, and the Externalizing and Internalizing scores obtained from different informants were always based on the same set of items. To ensure that the Total Behaviour Problem scores (consisting of the sum of the scores from all items on a checklist) obtained from different informants were based on the same set of items, items not present on all the checklists involved in a particular set of comparisons were omitted from the calculation of Total Behaviour Problem scores. Thus, in the comparisons of parent and child reports, the 16 items which were not present on both the CBCL and YSR were omitted from the calculation of Total Behavior Problem scores. Similarly, the 29 items not present on the TRF, CBCL and YSR were omitted from the calculation of Total Behavior Problem scores for the purpose of comparisons involving these three checklists.

To ensure that scores on the Externalizing and Internalizing Scales were always based on the same set of items, regardless of the age or sex of the child, or the informants being compared, the scores on these scales were all based on the items that make up the Externalizing and Internalizing Factors on the CBCL for 12-16 year old males. Once again, for each set of comparisons only items which were present on all the checklists involved in the comparisons were included in the calculation of scores on the two scales.

The modifications to the scoring procedures ensured that when comparisons focused on the number of problems reported by different informants, the scores compared were always based on the same set of items. However, the modifications were employed only for analyses which involved comparing the number of problems reported by different informants. For other analyses, such as the correlation between scores from different informants, the differences between the results obtained using the standard sets of items and the modified sets were minimal. In light of this, and to ensure comparability with other studies, the normal item content of the scales was retained with these analyses. It should be noted that the two month time period over which teachers were asked to base their ratings was not altered. This was to ensure that results from the teacher reports in this study could be compared with reports from other studies in Australia and overseas.

Cronbach's alpha (Cronbach, 1951) was employed to test the reliability of the modified scales. This statistic provides a measure of the internal or inter-item consistency of a scale and it represents the mean coefficient obtained from all possible split-half coefficients that are calculable for a given scale. Values of Cronbach's alpha range from 0 to 1 with higher values reflecting better inter-item consistency. Initially, the standard item content of the Total Behaviour Problem Scale, Externalizing Scale, and Internalizing Scale was retained and mean consistency coefficients were calculated for each scale using the scores obtained from parents, teachers, and children who participated in this study (Table 4.1). In the case of the Externalizing and Internalizing Scales, coefficients were based on scores from the CBCLs completed by parents of the 14-15 year old males. These coefficients provided the standard against which the reliability coefficients of the modified scales were to be compared.

The reliability coefficients obtained from the modified scales are shown in Tables 4.2 and 4.3. The coefficients shown in Table 4.2 were based on scores obtained using the 102 items present on both the CBCL and the YSR, and the set shown in Table 4.3 were based on scores using the 89 items present on CBCL, YSR and TRF. It can be seen that the reliability coefficients for the scales in both tables were high. In addition, there was little difference between the reliability coefficients for scales which included the full set of items (Table 4.1) and those in which some items had been omitted.

The distribution of scores obtained using the 102 items present on the CBCL and YSR are shown in the Appendix. There was little change in the distribution patterns when the thirteen additional items not present in the TRF were dropped and these distributions have not been included in the appendix. It can be seen that a number of the distributions are positively skewed, reflecting the fact that most children living in the community score within the normal range on the child behaviour checklists.

4.2 DATA ANALYSIS

Data analysis compared the scores obtained from the different informants on the Total Behavior Problem Scale, the Externalizing Scale, and the Internalizing Scale of the CBCL, TRF, and YSR. The Total Behavior Problem Scale provides an overall rating of children's problems while the Externalizing and Internalizing Scales provide more specific information about the nature of the problems. In view of the differences between narrow-band factors on the different checklists, scores from these factors were not used in the analysis of results.

The first set of analyses reported in this chapter compare the number of problems reported by different informants. These reports are genuine multivariate outcomes for a single observational unit (the child) which poses special problems for the testing of significances of differences between scores obtained from different informants. In addition, examination of the distribution of scores from the informants showed that they were positively skewed. Data analysis employed two different

approaches, both of which attempted to address aspects of the covariance structure implicit in this design and the distribution of the scores obtained from the informants. The first approach used a non-parametric statistic, the Wilcoxon Matched-Pairs Signed-Ranks Test, to determine the significance of differences between scores obtained from pairs of informants. The advantage of using this non-parametric statistic is that it makes few assumptions about the distribution of scores being compared (Gibbons, 1976). The disadvantage is that it cannot readily be extended to comparisons involving more than two groups. The second approach employed a standard (Normal Theory) analysis of variance which makes more assumptions about the distribution of scores in the groups compared, but which can include comparisons between more than two groups. The analyses of variance compared the size of the mean deviation of each informant's score from the average score for each child across all informants. This latter approach takes into account much of the covariance structure referred to above, without the loss of statistical power associated with multiple Signed-Ranks Tests. When consistent results are obtained using two different approaches to data analysis, it is less likely that the significances reported are an artifact of a particular statistical approach.

In order to retain the maximum number of items for each set of comparisons, results were analyzed separately for comparisons involving mothers and fathers (118 items), parents and children (102 items), and teachers, children and parents (89 items).

The second set of analyses reported in this chapter focuses on the degree of consistency between informants (Achenbach et al., 1987). As highlighted by Achenbach et al. (1987), clarification of the degree of consistency between informants is important because a very high consistency between certain informants "would mean that reports by one of them should serve as well as reports by more than one of them." (Achenbach et al., 1987). The Pearson product-moment correlation was chosen to describe the level of consistency of reports in this study because it has been widely used for this purpose (Achenbach et al., 1987, p. 214) and because it is a reasonably robust statistic provided that the distributions of scores being compared are not too skewed.

4.3 COMPARISON OF PARENT, CHILD AND TEACHER REPORTS DESCRIBING CHILDREN'S EMOTIONAL AND BEHAVIOURAL PROBLEMS

4.31 Comparison of Mother and Father Reports

The mean scores obtained from CBCLs completed by mothers and fathers are shown in Table 4.4. In three of the four groups, mothers consistently reported more problems than fathers on the Total Behavior Problem Scale, Externalizing Scale and the Internalizing Scale. Although the size of the differences between the mother and father scores were relatively small (representing less than 4% of the maximum score possible on individual scales), in the three groups, differences between mother and father scores on the Total Behavior Problem Scale reached a significance level of $p<0.05$. In the same three groups, differences between scores on the Internalizing Scale were significant at $p<0.05$ level. None of the

differences between the mother and father scores on the Externalizing Scale achieved this level of significance. This suggests that differences in the total number of problems reported by parents primarily reflected differences in their perception of internalizing problems. It should be noted that using the $p<0.05$ level of protection identified in Chapter 1, less than three differences would be expected to be significant by chance with the twelve comparisons.

4.32 Comparison of Parent and Child Reports

This section describes the results of comparisons between the number of problems reported by children and their parents. Initially, analyses tested whether there were significant differences between parent and child scores on the three broad band scales. Subsequently, the size of the differences between parent and child scores in groups containing the older and younger children were compared.

The mean scores from children and parents are shown in Table 4.5. In all four age and sex groups, the scores on the Total Behavior Problem Scale reported by the children were considerably higher than the scores reported by the parents. The child-reported scores were also higher than the parent-reported scores on both the Internalizing and Externalizing Scales. The size of the difference between parent and child scores was generally larger on the Internalizing Scale than on the Externalizing Scale primarily because the children, particularly the females, reported somewhat more internalizing problems than externalizing problems.

The Signed-Ranks Tests, testing for the statistical significance of differences between the scores obtained from parents and children revealed that the differences between the mothers and children, and between the fathers and children were highly significant ($p<0.001$) on all three scales in all the four age and sex groups. The mean deviation of each informant's score from the average score across all informants for each child are reported in Table 4.6. The results of the one way analyses of variance testing for the significance of differences between the mean deviations of informant scores were significant at $p<0.0001$ on all three scales in all four age and sex groups. All pairwise comparisons between mothers and children, and between fathers and children were significant at $p<0.001$.

The size of the differences between parent and child scores in the four age and sex groups are shown in Table 4.7. Analyses of variance were used to test whether the differences between parent and child scores varied significantly across the four age and sex groups. Where significant variations were identified, least significant difference contrasts (SAS Institute Inc., 1987) were used to compare the size of difference between scores obtained from parents and the older children with the difference between scores obtained from parents and the younger children.

It can be seen in Table 4.7 that the size of the difference between the mother and child scores varied significantly across the four age and sex groups on all three scales. For both the male and female children, least significant difference contrasts showed that the difference between the scores reported by the mothers and older children were significantly

greater ($p<0.05$) than the difference between the scores reported by mothers and younger children on all three scales. This was due to the combined effect of the older children reporting more problems than the younger children, while the mothers of the older children reported fewer problems than the mothers of the younger children (Table 4.5).

The mean difference between father and child scores in the four age and sex groups and the results of the analyses of variance testing for the significance of variations in these differences, are shown in Table 4.7. Although significant at the level of $p<0.05$, the variation across the groups was smaller than for mothers and children. On all three scales, the difference between the scores reported by the fathers and the older female children were significantly greater ($p<0.05$) than reported by the fathers and the younger female children. In contrast, the difference between the scores reported by the fathers and the older male children did not vary significantly from those of the fathers and the younger male children.

Finally, in each age and sex group, the size of the difference between the scores reported by mothers and children was compared with the size of the difference between the scores reported by the fathers and children. The t-test statistic was used for this comparison because the differences between parent and child scores showed a normal distribution. For both groups of younger children, and the 14-15 year old female children, there were significantly greater differences between the fathers and children than between the mothers and children on reports of internalizing problems (Table 4.8). This pattern reflected the differences between mothers and fathers described earlier where, except in the group

containing male 14-15 year olds, fathers reported that the children had fewer internalizing problems than were reported by the mothers.

4.33 Comparison of Parent, Child and Teacher Reports

This section describes the results of comparisons between the number of problems reported by teachers, parents, and children.

Scores on the scales used to compare the number of problems reported by teachers and the other informants were based on the 89 items that are present on the TRF, CBCL and YSR. As described earlier, for the purpose of these comparisons, the externalizing and internalizing scores were derived from the set of items which make up the externalizing and internalizing scales on the CBCL for 12-16 year old males, excluding those items not present on all three checklists.

The mean scores from teachers, parents, and children on the Total Behavior Problem Scale, Externalizing Scale, and the Internalizing Scale are shown in Table 4.9. It can be seen that on all three scales, teachers in the four age and sex groups consistently reported fewer problems than either the parents or the children. The differences between teachers and children were particularly pronounced. As well, the differences between teachers and female children were consistently greater than those between the teachers and the male children. This was particularly evident for reports of internalizing problems. This latter pattern reflected the combined effect of teachers reporting somewhat fewer internalizing problems for female than male children, while the female children themselves reported having more internalizing problems than the males.

All the differences between teacher and mother scores, and teacher and child scores were significant at the level of $p<0.0001$ (Signed-Ranks Tests). Differences between the teacher and father scores were all significant at $p<0.01$. The mean deviation of each informant's score from the average for each child across all informants are shown in Table 4.10. It can be seen that on all three scales in all four age and sex groups, the differences between the mean deviations were significant at the level of $p<0.0001$. For scores from all three behaviour scales, least significant difference contrasts showed that differences between teachers and mothers, and teachers and children were all significant at the level of $p<0.0001$. Differences between teachers and fathers were all significant at the level of $p<0.01$.

The mean difference between the scores reported by teachers and those reported by children in each of the four age and sex groups are also shown in Table 4.11. Analyses of variance showed that the difference between the scores reported by teachers and children varied significantly (Table 4.11) across the groups on all three scales. Least significant difference contrasts were again used to compare the size of the difference between teacher and child scores in the groups containing the older children with the size of the difference in the groups containing the younger children. These contrasts showed that difference between scores reported by teachers and the older male children were significantly ($p<0.05$) greater than the difference between teachers and the younger male children on the Total Behavior Problem Scale and the Externalizing Scale. The differences between the groups containing the older and younger female groups were not significant at the level of $p<0.05$.

The mean difference between teacher and parent scores in the four age and sex groups are shown in Table 4.11. In contrast to the findings with children and teachers, the size of the difference between scores reported by teachers and parents varied little across the four age and sex groups and, with one exception, none were statistically significant.

A plausible reason for the differences in size of the parent, child and teacher scores was that they arose because of differences in the response patterns of the informants, rather than because of differences in the number of items endorsed. For example, it was possible that parents and children had endorsed the same number of items, however, parents had endorsed the items as "Somewhat or sometimes true" while children endorsed them as "Very true or often true". If true, this would have resulted in large differences between the parent and child scores, even though there was no difference in the number of items endorsed.

Initially, the percentage of items scored as "1" and "2" on each scale by the parents, children and teachers were examined (Table 4.12). It can be seen that the percentage of items endorsed by the different informants as "1" or "2" appears to be similar. In order to test more formally the possibility that different response patterns were making a major contribution to differences in the size of scores obtained from different informants, the mean parent and child scores were re-calculated after the scores "1" and "2" for each item had been collapsed to a single category of "1". The sum total of these new scores represented the number of items endorsed by each informant.

The mean number of items endorsed by each informant is shown in Table 4.13 and the results of one way analyses of variance testing for the significance of differences in the mean deviation of each informants score from the average across all informants are shown in Table 4.14. It can be seen that the pattern of differences and significance levels were unchanged from those reported earlier in this section. There were still highly significant differences between the reports from children, parents, and teachers. In addition, the least significant difference contrasts between parents and children, and teachers and children were all highly significant ($p<0.0001$). The significance levels for the differences between mother and father scores were the same as those reported in Section 4.31. These results suggest that the differences between the informants reflected the total number of items endorsed, rather than being solely due to differences in the scores on individual items reported by different informants.

4.34 Estimated Correlations between Scores from Mothers, Fathers, Children, and Teachers

Estimated Pearson Correlations were calculated for the scores obtained from the mothers, fathers, children, and teachers on the Total Behavior Problem Scale, Externalizing Scale, and the Internalizing Scale. There were minimal differences between correlations based on scores obtained from scales containing the standard items, and scores from the scales from which some items had been omitted. In view of this, and to ensure comparability with previous studies, the correlations reported in this study are those obtained using scales comprised of the standard items specific to each checklist.

The estimated Pearson Correlations for the total problem scores obtained from mothers, fathers, children, and teachers are shown in Table 4.15 and Table 4.16. In all four age and sex groups, the highest correlations were between the scores obtained from parents, and the lowest were between scores from children and teachers. Correlations between children and parents generally occupied an intermediate position. This pattern was also evident for correlations between scores on both the Externalizing and Internalizing Scales. (Tables 4.17 to 4.20).

In general, the difference in the size of the correlations between the externalizing and internalizing scores from the same pair of informants was small. Although the correlation between the externalizing scores was generally larger than the correlation between the internalizing scores, the only differences which achieved a $p<0.05$ level of significance (using an r to z transformation to test for the significance of differences) were those between scores describing the older female children reported by teachers and mothers, and teachers and fathers. In each case, the correlation between the externalizing scores was significantly higher than the correlation between the internalizing scores reported by these informants.

4.4 COMPARISON OF PARENT AND CHILD REPORTS DESCRIBING CHILDREN'S COMPETENCIES

An advantage of the child behaviour checklists is that they include reports of children's competencies as well as their behaviour problems. This section describes comparisons of reports describing children's competencies, obtained from parents, children and teachers.

4.41 Data Analysis

Although the competencies identified on the CBCL and YSR have similar names, there are important differences in the way scores are calculated for the two checklists. For example, items on the CBCL describing special class placement, failing a grade, and other school problems, are not present on the YSR. In light of this, a separate school performance scale is not included on the YSR although self-rated scores of performance in academic subjects do contribute to the Total Competence score on the YSR (Achenbach & Edelbrock, 1987, p. 10). The differences between the CBCL and YSR competence items result in differences in the possible range of scores on the two questionnaires for the Total Competence score. On the CBCL the range of possible scores is 0-30 while on the YSR the range is 0-27.

The TRF focuses particularly on childhood competencies which can be identified by teachers. As a result, a somewhat different set of competencies is scored on this checklist than on the YSR and CBCL. Competencies on the TRF are labelled school performance, working hard, behaving appropriately, learning, and happy. The overall rating of competence on the TRF is based on the sum of the latter four scores and is called "Total Adaptive".

In view of the differences between the scoring procedures on the three checklists, comparisons of scores from different informants are restricted to the use of Signed-Ranks Tests to test for the significance of differences between competence scores on reports obtained from parents and children. In addition, estimated Pearson correlations are reported

for Total Competence/Total Adaptive scores on the CBCL, YSR and TRF. As described earlier, pilot studies showed that some 10-11 year olds were not able to complete independently the competence items on the YSR and this section was not included in questionnaires for the 10-11 year old age group.

4.42 Results

Mean competence scores obtained from the different informants are shown in Table 4.21. The male and female children consistently reported lower scores than their parents on the Total Competence Scale but somewhat higher scores on the Activities and the Social Scales. The lower child scores on the Total Competence score were largely due to the influence of lower ratings of academic performance reported by children. Differences between reports obtained from mothers and fathers were generally small.

The results of the Signed-Ranks Tests testing the significance of the differences between parent and child scores are shown in Table 4.22. Differences on the Total Competence Scale between mothers and children, and between fathers and children in both groups were significant at the $p<0.001$ level. This analysis was repeated after items not present on both the CBCL and YSR were omitted. The omission of items caused minimal changes to the mean scores on the Total Competence Scales on the CBCL (<0.5 on any affected score), and had no effect on the significance levels in Table 4.22. Although the magnitude of the differences between the competence scores reported by mothers and fathers were small, five differences reached a significance level of 0.05. If the $p<0.05$ level of protection identified in Chapter 1 is employed, less than three

differences would be expected to be significant by chance with the twelve comparisons.

Estimated Pearson correlations are shown in Table 4.23 and Table 24. The pattern is very similar to that found with the behaviour problem scores. The correlations between competence scores reported by parents are the highest in each age and sex group, correlations between scores reported by parents and children are intermediate, and the correlations between teacher and child scores are the lowest. It should be emphasized, however, that items comprising teacher scores are different from those in the CBCL and YSR.

4.5 DISCUSSION

This chapter examined agreement amongst the informants traditionally used to identify childhood emotional and behavioural problems. A major strength of the study is that it obtained separate reports from mothers, fathers, teachers, and children, describing children in two different age groups. This made it possible to compare reports describing children of different gender and age living in the community.

The principal findings are that the children reported significantly more externalizing and internalizing problems than their parents. As well, both the children and parents reported significantly more externalizing and internalizing problems than the teachers. Finally, the difference between the number of problems reported by the older children and their mothers was significantly greater than the difference between the number of problems reported by the younger children and their mothers. This

pattern was also evident for fathers and the female children, however, the variation across the age groups was not evident with fathers of the male children.

Although there were some differences between the results found with the older and younger children, the overall pattern of results obtained from the different age and sex groups was very consistent. Children of both sexes and in both age groups reported, on average, many more problems than the mothers, fathers, and teachers. For example, the older female children reported almost twice as many problems as their parents, and three times as many problems as their teachers. Differences between the mothers and fathers, although statistically significant, were much smaller. Finally, there was a consistent pattern in all the age and sex groups for teachers to report the least number of problems.

Assessing the clinical significance of the differences identified between the parent and child scores is difficult. One approach which can be used is to compare the size of the differences between the mean parent and child scores identified in this study, with the size of the differences of the mean CBCL or YSR scores across the community and clinic-referred groups in the normative studies conducted by Achenbach and Edelbrock (1983, 1987). This is a similar approach to that used by Sawyer et al. (1988) in order to identify the clinical significance of the differences between reports from parents and children when describing family psychological adjustment. The differences between the total problem scores reported by the parents and the 14-15 year old children were of a similar magnitude to the differences between the YSR scores across the community and clinic-referred groups, and ranged from 40-50% of the differences between CBCL scores across the community and clinic-referred

groups (It should be noted that the size of the differences between the mean YSR scores across the community and clinic-referred groups in the studies by Achenbach and Edelbrock were smaller than the differences between the mean CBCL scores across the same groups). The differences between the total problem scores reported by the parents and 10-11 year old children were somewhat smaller, representing 20-30% of the differences between scores across the community and clinic-referred groups. These comparisons suggest that the size of the differences between parent and child scores are clinically significant and that a clinician's assessment of a child's behaviour may vary depending on the relative weight given to reports from different family members. However, it is important to note that this approach assumes that the differences between the mean CBCL or YSR scores across the community and clinic-referred groups in the studies by Achenbach and Edelbrock (1983, 1987), represent a clinically significant difference. Possible limitations of using the difference between community and clinic-referred groups to assess clinical significance are discussed in Chapter 7.

Although the size of the correlations between scores from different dyads varied, the pattern was similar in each age and sex group. In particular, the highest correlations were between the scores reported by the mothers and fathers and the lowest were between scores from the children and teachers. This pattern was evident for ratings of both externalizing and internalizing problems. The consistency of results in the different groups suggests that the findings reflect general patterns which will be found in reports describing emotional and behavioural problems in older children of both sexes. The varying levels of parent-child agreement reported by Edelbrock et al. (1986) for children of different ages were not present in this study. It is possible that this

was due to the much narrower age range of children in this study as compared to the age range of children in the study by Edelbrock et al (1986).

The finding that children reported more problems than their parents is consistent with the results from an earlier study by Achenbach and Edelbrock (1987, p. 112). The study by Achenbach and Edelbrock (1987) also reported that the number of problems identified by mothers of female children was greater than the number identified by fathers, a finding which was replicated in this study. Results in this study suggest that the reason for the difference between the number of problems reported by mothers and fathers was primarily that for female children, mothers reported more internalizing problems than fathers.

There are several plausible reasons why children may report more problems than other informants. For example, it is possible that children may have a lower threshold than parents for reporting problems. This possibility was discussed by Kashani et al. (1985) when they pointed out that children and parents may differ in their judgement as to when it is warranted to report a problem. Although it was not possible to specifically test this hypothesis, it was possible to test whether the difference between the child and parent scores in this study arose primarily because children reported more problems, or whether the children reported the same number of problems but endorsed individual problems as being more severe (resulting in a higher score on the behaviour scales). These analyses suggested that the higher scores reported by children were due to the children reporting more problems, rather than because they endorsed individual problems as being more severe. Similarly, the lower teacher scores appeared to be due to

teachers reporting fewer problems rather than because they reported the same number of problems, but rated individual problems as less severe.

Another possible reason for children reporting more problems is that they draw on knowledge of their behaviour in a much wider range of settings than the other informants. For example, it is likely that the reports from parents and teachers largely reflect their perception of a child's behaviour in the context of their own relationship with the child. In contrast, when children rate their problems, they may be describing the sum total of their problems in a large number of different settings. Finally, as highlighted by many investigators, some problems, particularly those involving internalizing symptoms, may not be apparent to an external observer. This might explain why children reported more internalizing problems than the other informants.

Although previous studies have found that more internalizing problems were reported by children than parents, in these studies more externalizing problems were reported by parents (Herjanic & Reich, 1982; Edelbrock et al., 1986). In this study both internalizing and externalizing problems were reported more frequently by children. One possible explanation for the different results is that the children in this study were selected from the general community rather than from amongst clinic-referred children. This explanation is supported by the results of a study by Offord, Adler, and Boyle (1986) who also found that children and adolescents consistently reported more antisocial symptoms than their parents. It is possible that clinic-referred children represent a special group in which parents report more externalizing problems than their children. In contrast, the more general pattern amongst children living in the community may be for the children to

report more externalizing and internalizing problems than their parents. This issue is examined in greater detail in Chapter 7.

One study which did focus on a community population of children was that of Verhulst et al. (1987). In contrast to the results found in this study, Verhulst et al. (1987) reported that most disagreements about the number of the children's problems were due " ... to higher parent scores. Fears was the only content area in the total sample on which children scored higher than parents" (p. 464). There are several possible explanations for the different results in the study by Verhulst et al (1987). Firstly, Verhulst et al. (1987) employed semi-structured interviews rather than checklists to collect information from parents and children. In addition, different semi-structured interviews were employed with the parents and children. The Child Assessment Schedule (Hodges et al., 1982) was used with parents and the semi-structured interview developed by Graham and Rutter (1968) was employed with children. Finally, "the parent" interviewed in the study by Verhulst et al. (1987) varied, including mothers, fathers, and joint interviews with both parents.

As noted by Verhulst et al. (1987), use of a semi-structured interview may provide the opportunity for raters to establish better rapport with children. For example, Verhulst et al. (1987) suggested that semi-structured interviews may enable raters to explain more clearly to children the nature of the symptoms about which they are being asked. This clearer explanation of symptoms may also lead to fewer endorsements by the children, and possibly more endorsements by parents. However, it is also possible that the presence of an interviewer intimidates children and this results in them reporting fewer problems. For example, some of

the children who did not wish to be interviewed in this study said that they would be willing to participate in further mail surveys because they found them less intimidating than being interviewed directly.

The finding that teachers reported considerably fewer problems than the other informants is important. It is also consistent with the results of a study by Verhulst and Akkerhuis (1989) which used the child behaviour checklists to obtain reports from parents and teachers. When clinicians assess children it is common practice to seek advice from teachers in order to try to establish the severity of a child's problems. If a teacher rates a child's problems as less severe, this may have a significant influence on the conclusions a clinician will draw about the severity of a child's difficulties, and the type of treatment plan that the clinician will initiate to help the child.

In contrast to the findings in this study, the earlier study by Offord et al. (1986) reported that teachers of primary school children described more antisocial problems than parents. The response rate from teachers of the older children in the study by Offord et al. (1986) was too low for their questionnaires to be used for similar comparisons. There are several possible explanations for the different results obtained from teachers in this study and the teachers who participated in the study conducted by Offord et al. (1986) in Ontario. Firstly, the study by Offord et al. (1986) was reporting on younger children and it is possible that teachers of these children show a different reporting pattern than teachers of older children. Secondly, as mentioned earlier, the response rate for teachers in the present study was relatively low and it is possible that children with more problems at school are under-represented in the study sample. Finally, it is possible that opportunities for

teachers to observe children's problems may differ in South Australia and Ontario. For example, it is possible that schools in South Australia are somewhat more regimented and, except when problems are severe, teachers have less opportunity to observe childhood problems.

Subsequent discussions with teachers suggested two further explanations for the relatively low frequency of problems reported by teachers in this study. Teachers suggested that, in South Australia at least, some teachers may be unwilling to acknowledge problems amongst students in their own classroom. This is because they are concerned that problems amongst their students may reflect badly on their ability as a teacher. In addition, they suggested that if teachers either didn't understand the origin of a problem or if they felt that it arose because of stresses external to the school, they may tend not to report the problem. The latter is very similar to the comment made by Garrison and Earls (1985) who suggested that teachers are reluctant to identify childhood symptoms which lie outside the school context. Touliatos and Lindholm (1981) have also suggested that parents may report more problems than teachers because they are less tolerant of their children, expect more of them, and have less knowledge about normal child development. Finally, it is possible that teachers reported fewer problems than the other informants because the Teacher Report Form requires teachers to report on children's behaviour over a two month period. The Child Behavior Checklist and Youth Self-Report Form ask informants to report on children's behaviour over a six month period. It should be noted that, despite this restriction, the study took place at the end of the school year and teachers would have known the participating children for at least eight months.

The correlation between the scores from informants in this study were generally higher than those reported in previous studies (Achenbach et al., 1987). Interestingly, the study by Verhulst et al. (1987) which focused on children in the community also found somewhat higher correlations than have been described in earlier studies. It is possible that the consistency between informant reports in community populations may be higher than that which applies with clinic-referred children who have been the focus of most earlier studies. It is also possible that the higher correlations found in this study reflected the fact that the study used very similar questionnaires with all the informants. In previous studies the use of different questionnaires with different informants may have adversely affected the consistency found between reports from different informants. This result has clinical relevance because it suggests that if clinicians are aware of the different reporting patterns of different informants, they may be able to adapt their clinical assessment to take into account this phenomenon.

In summary, the results reported in this chapter identify significant differences between the number of problems reported by parents, children, and teachers. These results have important implications for the psychiatric assessment of children. For example, different diagnostic categories in the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) (American Psychiatric Association, 1980) are defined on the basis of the number of symptoms present in a child. At the threshold of a DSM-III diagnosis, the presence or absence of a single symptom can influence whether or not a child is diagnosed as having a psychiatric disorder. The large difference between the number of problems identified by informants in this study suggests that clinicians would likely have problems if they attempted to achieve consensus between the different

informants about the severity of a child's problems. This is a significant problem because, as noted by Edelbrock and Costello (1988), although the DSM-III identifies what to assess, it provides little guidance about how to assess it. Currently there are no generally accepted guidelines which can be employed to resolve disagreements between informants when attempting to establish a child's diagnosis.

CHAPTER 5. COMPARISON OF THE NUMBER OF "CASES" AND ESTIMATED PREVALENCES IDENTIFIED FROM PARENT, CHILD, AND TEACHER REPORTS

5.0 INTRODUCTION

To date, few studies have investigated the level of agreement which exists between informants when a categorical approach is used to classify childhood emotional and behavioural problems. This is a significant omission because clinicians often employ this approach when assessing and treating childhood problems. As well, when categorical approaches are used to classify childhood problems, attention is focused not only on the number of problems identified by different informants, but also on the type of problems which are identified.

This chapter reports the level of agreement between parents, children, and teachers when three different categorical approaches were used to classify the children's emotional and behavioural problems. The first categorical approach employed the cutoff-scores (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986; Achenbach & Edelbrock, 1987) on the CBCL, TRF, and YSR completed by parents, teachers, and children, to classify children as "cases" or "non-cases". The classification was done independently according to each child's scores on each checklist. The level of agreement between the different informants about the groups to which the children were assigned and the frequency with which children were assigned to the different groups is reported.

The second approach employed the Child Behavior Profiles (Achenbach & Edelbrock, 1983), which can be scored from CBCLs completed by parents, in order to classify children's problems. In contrast to cutoff scores which apply to individual scales, the Child Behavior Profiles classify children's problems on the basis of the entire pattern of the child's CBCL profile and the magnitude of its scores (see Section 5.22). As such, it seemed plausible that agreement between parents might be different when comparisons were based on agreement about a child's profile/s, as compared to agreement based only on a single scale score. The Child Behavior Profiles can only be scored from the CBCL, hence it was not possible to include reports from children and teachers in this comparison.

The final categorical approach compared the individual behaviour problem item scores which were reported by parents, children, and teachers. The level of agreement between different informants and the percentage of informants reporting each item are described. The latter analysis is important because it provides information about the type of problems identified by different informants and which problems would be under-reported if assessments employed only one or two informants.

5.1 COMPARISON OF CASES

The analyses in this section concentrate on three questions:

1. What was the prevalence of cases when different informants were used to obtain information about children's behaviour?

2. What percentage of the children who were identified as cases when information was obtained from all four informants, would be identified if information was obtained from only one, two, or three of the informants?

3. What percentage of children were identified as cases by only one informant?

5.11 Subjects

The subjects for this comparison were the 336 children for whom completed questionnaires were available from two parents, a teacher, and the child. This represented 75% of the children living in two-parent families who participated in the study. The loss of subjects was primarily due to the absence of a teacher questionnaire for 95 children, and a father questionnaire for 55 children in the two-parent families.

In 57 families there was a single parent and the children were not eligible for inclusion in the analyses described in this chapter. The percentage of single-parent families in the total study sample was 7% in the highest socio-economic school group, 11% in the middle socio-economic school group, and 19% in the lowest socio-economic school group. As a result, the sample of children included in this chapter is somewhat under-representative of children in the lowest socio-economic school group.

The mean scores from mothers, fathers, children, and teachers for the 336 children for whom completed checklists were available from all informants

are shown in Table 5.1. The differences between the scores for the children in two-parent families included in this chapter and those excluded because one checklist was not available were minimal and none were statistically significant. The principal reason that scores for children in these analyses were slightly lower than for the entire sample of children was that children from single-parent families, who were not included, tended to have more problems than the children from two-parent families.

5.12 Data Analysis

Cutoff scores recommended for use with the Total Behavior Problem Scale, Externalizing Scale and the Internalizing Scale taken from the parent, teacher and youth checklists (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986; Achenbach & Edelbrock, 1987) were utilized to define children as cases or non-cases. Achenbach and Edelbrock (1983) state that these cutoff scores enable users of the checklists "to discriminate in a more categorical fashion between children likely to resemble our clinical sample and those more likely to resemble our nonclinical sample." (p. 62). Separate cutoff scores are provided for the Total Behavior Problem Scale, Externalizing Scale and Internalizing Scale for children of different age and sex on the three checklists. In general, these scores are equivalent to the 90th percentile of the scores on the CBCL, and the 89th percentile for scores on the YSR and TRF. Although Hensley (1988) has suggested tentative cutoff scores for use with Australian populations, these are still to be validated and the cutoff scores recommended by Achenbach and Edelbrock (1983, 1986, 1987) have been employed in this chapter.

For the purpose of validation studies, "referral for mental health services" was chosen as the criterion against which to test the criterion validity of the scales on the CBCL (Achenbach & Edelbrock, 1983, p. 55). Achenbach and Edelbrock (1983) reported that " ... Using referral as a criterion and the 50-50 split of referred versus nonreferred in our samples, the total misclassification rate was 17.85% (9.85% of nonreferred children scoring in the clinical range and 25.9% of referred children scoring in the normal range)." (p. 65) On the TRF, the recommended cutoff score for the Total Behavior Problem Scale classified 88.8% of nonreferred children and 44.4% of referred children in the normal range, an overall misclassification rate of 27.8% (Achenbach & Edelbrock, 1986, p. 55). Finally, on the YSR, the cutoff on the Total Behavior Problem scale classified 89.5% of nonreferred children and 65.0% of referred children in the normal range, an overall misclassification rate of 37.8% (Achenbach & Edelbrock, 1987, p. 53). On each questionnaire, somewhat higher misclassification rates were found with cutoff scores on the Externalizing and Internalizing Scales than applied on the Total Behavior Problem Scale.

Achenbach and Edelbrock (1987) noted that misclassification rates in populations containing proportions of adolescents differing from that used in their validation studies (50% referred and 50% not referred) will vary. For example, if it is assumed that approximately 90% of adolescents in the community are not disordered, the approach recommended by Achenbach and Edelbrock (1983, 1986, 1987) can be used to calculate misclassification rates that would apply for adolescents in the community. In this case the misclassification rates for the three checklists would be 11.4% on the CBCL, 14.5% on the TRF and 16.5% on the YSR.

Two different statistical approaches were used to investigate the level of agreement between informants when scores from their respective child behaviour checklists were used to classify children as cases or non-cases. In the first approach, the percentages of children actually identified by two or three informants were identified. These were then compared with the percentages of children it was estimated that two or three informants would identify, assuming that identification by one informant was totally independent (in the strict probabilistic sense) of identification by any other informant. These estimated percentages (P_E) were based on the percentages identified by individual informants as follows:

$$P_{E,2} = P_1+P_2-(P_1 \times P_2)$$

$$P_{E,3} = P_1+P_2+P_3-(P_1 \times P_2)-(P_1 \times P_3)-(P_2 \times P_3)+(P_1 \times P_2 \times P_3)$$

P_1 = Actual percentage from informant1

P_2 = Actual percentage from informant2

P_3 = Actual percentage from informant3

P_E = Estimated percentages from two ($P_{E,2}$) or three ($P_{E,3}$) informants

P_A = Actual percentage from two or three informants

It would be expected that P_E would be greater than P_A if there was a high level of agreement between informants about whether or not a child was a case.

The second approach employed the kappa statistic to measure the level of agreement between different informants. The kappa statistic (Cohen, 1960) varies from negative values for less than chance agreement, through 0 for chance agreement, to 1.0 for perfect agreement. Different criteria have been suggested with regard to what is an acceptable level of kappa. In their study of agreement between mothers and children, Herjanic and Reich (1982) defined kappas of 0.50 and over as "high", 0.30 to 0.49 as "middle", and less than 0.29 as "low". Costello et al. (1984) used somewhat stricter criteria and defined 0.60 or higher as "high", 0.40 to 0.59 as "moderate", and 0.20 to 0.39 as "low". Kappas below 0.20 were considered by Costello et al. (1984) to represent non-agreement. Results in this study are reported using the criteria employed by Costello et al (1984). Herjanic and Reich (1982) also suggested that kappas are not useful for very low frequencies and responses. In light of this, kappas are not reported where fewer than 10 cases were identified by the two informants being compared. (A similar argument holds for situations where very high numbers of cases are identified by informants, however, this situation did not apply in these analyses.)

As highlighted by Herjanic and Reich (1982), "If there were sufficient positive responses by [the informants] to a particular symptom, the result could show middle to high agreement, yielding a kappa above .30, but could still be called 'asymmetrical', because [one informant]

reported significantly more positive responses than the other." (p. 311). This pattern was called "asymmetrical reporting" by Herjanic and Reich (1982, p. 311). In this study, the McNemar statistic (Armitage & Berry, 1987) was used to identify asymmetrical reporting and visual inspection of results revealed which informant was responsible for the asymmetry.

5.13 Results

1. What prevalences are identified when reports are obtained by different informants?

Estimates of prevalences which would apply if reports were obtained from different individual informants are shown in Table 5.2. These estimates were calculated with the weighting appropriate to the stratified sampling scheme employed in the study. Large differences can be seen in the prevalence estimates that were found when reports were obtained from different informants. For example, when cutoff scores on the Total Behavior Problem Scale were used, prevalence estimates for male 10-11 year olds would range from 4.7±2.3 to 17.9±4.1. In general, the highest estimates were obtained when reports were obtained from mothers and the lowest estimates when reports were obtained from children.

One apparent inconsistency revealed in the prevalence estimates is that although the mean scores reported by children were higher than those reported by parents and teachers, the higher mean scores were not reflected in higher percentages of cases identified by the child informant. The principal explanation for this apparent anomaly is that the recommended cutoff scores for the YSR are considerably higher than the cutoff scores for the CBCL and TRF. As a result, although the mean

child scores were higher than that of other informants, this was not immediately reflected in the prevalences identified by the children. It should also be recalled that the cutoff scores were based on samples of children in the U.S.A. Results in this study suggested that these latter children may report somewhat more problems than children in Australia. If this pattern also applied to clinic-referred children, it would suggest that different cutoff scores may be more appropriate for Australian children. At present there is little information available about the YSR scores which apply for clinic-referred children in Australia.

2. What percentage of disordered children are identified if one, two, or three informants are used to collect information about the children?

The total number of children identified as cases by at least one informant from scores on the Total Behavior Problem Scale, Externalizing Scale and the Internalizing Scale are shown in Table 5.3. For example, it can be seen that a total of twenty-one 10-11 year old male children were identified as cases using the cutoff score from the Total Behavior Problem Scale, sixteen 10-11 year old male children were identified using the cutoff score on the Externalizing Scale and twenty-one 10-11 year old male children were identified using the cutoff score on the Internalizing Scale. As it was not essential for the purpose of comparisons described in this section that cases be classified as exclusively Internalizers or Externalizers, the more stringent criteria suggested by Achenbach and Edelbrock (Achenbach & Edelbrock, 1983, p. 34) to achieve this level of classification were not employed.

The percentage of the total number of cases identified by different informants using scores from the Total Behavior Problem Scale are shown in Table 5.4. The percentages identified by different informants were very similar for children in each of the four age and sex groups. In each group, the greatest number of cases were identified if mothers were used as the informant, with smaller percentages identified if fathers, teachers or children were used as the informant. It is important to note, however, that even mothers identified only 66% of the cases who would have been identified if all four informants are used.

The cutoff scores on the Externalizing and Internalizing Scales were employed to identify whether the percentage of cases identified by different informants varied, depending on the type of problems being rated. The percentage of children identified as cases by different informants on these scales are shown in Tables 5.5 and 5.6. Overall, the percentage of cases identified by different informants on the two scales was similar, but on both scales there were important differences between the age and sex groups. For example on the Internalizing Scale, in the group containing the older female children, the largest percentage of cases (50%) were identified when the child was used as the informant. In contrast, on the same scale in the younger male group, the largest percentage of cases (76%) were identified if mothers were used as the informant. Similarly, on the Externalizing Scale the largest percentage of cases in the older male group (72%) were identified by fathers but in the younger male group the largest percentage (63%) were identified by teachers.

On all the Scales, a larger percentage of cases were identified when two or three informants were used to identify cases. Finally, the relative

percentage of cases identified by different combinations of informants on the Externalizing and Internalizing Scales was similar. For example, when mothers and fathers were both used to obtain information on the Externalizing Scale, 78% of cases were identified, while these informants identified 74% of cases on the Internalizing Scale.

The percentage of cases (based on scores on the Total Behavior Problem Scale) which it was estimated that two or three informants would identify (P_E) are shown in the right hand column in Table 5.4. In this estimation it was assumed that for any one child, identification as a case by one informant was totally independent of identification by any other informant. It might be suspected that if a child was identified as a case by one informant, there would be a much higher chance of being similarly classified by a second informant. However, the similarity of the estimated percentages (P_E) and the percentages actually (P_A) found, suggests that the proportion of children who were simultaneously identified as "cases" by both informants was no larger than might be expected by chance. Consistent with these findings, results obtained when the kappa statistic was used to measure the level of agreement between informants (Tables 5.7-5.8) suggested that agreement with regard to the identification of cases was low. In many instances, the kappas were within the range defined as non-agreement.

On both the Externalizing and Internalizing Scales there was also little difference between the percentages of cases actually identified (P_A) and the estimated percentages (P_E) calculated on the assumption that informants were rating independently of each other (Table 5.5-5.6). This suggests that for cases defined on the basis of either externalizing or internalizing problems, informants were rating independently of each

other. Once again, the same result was obtained when the kappa statistic was used to evaluate the level of agreement between different informants (Tables 5.7-5.8).

It has been suggested that an advantage of the kappa statistic is that it takes into account the extent that agreement between informants exceeds chance expectations. In particular, it has been considered that the statistic takes into account the high level of agreement that is found between informants when "symptoms are absent or rare" (Herjanic & Reich, 1982, p. 310). This is a very important consideration in epidemiological studies of childhood psychopathology.

In order to test the influence on the kappa statistic of large numbers of agreements that a child is not a case, or does not have symptoms, the kappa statistics described earlier were re-calculated with all 336 children included in the calculation. This made it possible to examine the effect on the kappa statistic of including a large number of agreements that a child is not a case. In the earlier calculations the kappa statistics (reported in Tables 5.7 and 5.8) were based on only those children who had been identified as a case by at least one informant. That is, the kappa statistic was used to provide a measure of how much agreement there was between informants describing the sub-sample of children who had been identified as cases by at least one informant. In the calculations which included all 336 children, the total number of cases identified by one or both of the informants obviously remains the same, however, a much larger percentage of the agreements between informants are now based on agreement that a child is not a case. (see, for example, Table 5.9). The results of the calculations which included all 336 children are shown in Tables 5.10 and 5.11. It can be seen that

much higher kappas were obtained when the results from all children were included in the calculation of the kappa statistic. It appears that the increased size of these kappas were due to influence of the large number of agreements between parents that a child was not a case. This finding suggests that caution is necessary in the interpretation of the meaning of kappa when the prevalence of cases or symptoms being investigated is low.

3. What percentage of the disordered children are identified by only one informant?

The percentage of children identified by only one informant as falling in the clinical range on each of the three broad band scales is shown in Tables 5.12 to 5.14. Once again, there is considerable consistency in results obtained from the four different groups. In particular, approximately 50% of male and female children identified as cases in both age groups are identified by only one informant. In the younger age group mothers and fathers identified approximately equal numbers of children, however, in the older age group fewer children were identified solely by fathers.

5.2 COMPARISON OF CHILD BEHAVIOUR PROFILES

Results in this section compare the Child Behavior Profiles scored from CBCLs completed by mothers with those scored from CBCLs completed by fathers.

The Child Behavior Profiles (Achenbach & Edelbrock, 1983) provide a means by which childhood disorders can be classified into different groups. To date, profiles can only be scored from CBCLs and the development and use of the profiles has focused primarily on children attending mental health clinics. In this context, the profiles provide a quantitative method of discriminating between children who differ in their behaviour disorders. It is important to note that the Child Behavior Profiles are distinguished by the entire pattern and magnitude of a child's profile score on the CBCL, rather than by their score on a single scale.

5.21 Subjects

Subjects used for this comparison were children for whom a completed CBCL was obtained from both their mother and father. In addition, because the profiles were designed to discriminate between children who differ in their type of behaviour disorders, rather than to distinguish between children with and without behaviour disorders, only profiles from children whose score on the Total Behavior Problem Scale exceeded the recommended cutoff score on at least one CBCL were included.

5.22 Data Analysis

The standard profiles for each age and sex group were originally constructed (Achenbach & Edelbrock, 1983) by means of centroid cluster analyses using CBCL scores from checklists which had been completed by parents of clinically-referred children. Two samples of 250 clinically-referred children of each sex and in each age range (6-11 and 12-16) were used to construct the profiles. Initially, children within one age and sex group were grouped according to the similarities between their

profile patterns measured by means of one-way intraclass correlations between profiles. Children "... whose profiles showed the highest ICCs with each other were grouped together into clusters. The *centroid* - a profile formed by averaging all the profiles in the cluster - constitutes the operational definition of each cluster." (McConaughy, Achenbach, & Gent, 1988, p. 487). The second sample of children in each age and sex group were then cluster-analyzed. The approach of employing two separate samples was used to guard against the possibility that the clusters identified in one sample were, by chance, peculiar to a particular sample. The intraclass correlations between the centroids obtained in the two samples were then calculated. "Each cluster whose centroid was found to correlate significantly ($p<0.05$) with a centroid in the other sample was considered to be reliable. The centroids of the two samples for a particular sex/age group were averaged to yield the final operational definition of each cluster." (Achenbach & Edelbrock, 1983, p. 76). As can be seen in Tables 5.15 to 5.18, either six or seven profiles can be identified from CBCL scores for each age and sex group. For classifying individual children, the ICC is calculated between the child's profile and the centroids of each of the clusters. The child is classified according to profile type with which the child's profile most highly correlates.

Users of the CBCL can stipulate the intraclass correlation required for a child to be considered a member of a particular profile group. Higher correlations are associated with greater homogeneity within the groups at the cost of a narrower coverage of children. Lower correlations are associated with less homogeneity allowing for the inclusion of a greater range of children. For the purpose of this study a profile had to show an intraclass correlation of ≥ 0.35 with one of the standard profiles,

and a score on the Total Behaviour Problem Scale between 25 and 100 inclusive, for the child to be considered a member of the profile group. This standard is the same as that used by McConaughy et al. (1988) in a recent study testing the construct validity of the Child Behavior Profiles.

5.23 Results

The comparison of profiles obtained from scores on the CBCLs completed by mothers and fathers in the four groups is shown in Tables 5.15 to 5.18. It can be seen that the number of profiles varies considerably in the four groups. For example, in the group consisting of 14-15 year old female children relatively few profiles met the criteria for inclusion in the comparison. In contrast, considerably more profiles were identified from the group containing the younger female children. In view of the fact that the number of profiles and children was small, no statistical analysis has been attempted with these results.

For the purpose of this comparison, three levels of agreement were defined. Full agreement was defined as both parents agreeing about the child's profile type and both scoring the child in the clinical range on the Total Behavior Problem Scale. Partial agreement was defined as both parents agreeing about the child's profile type but only one parent scoring the child in the clinical range. Non-agreement was defined as parents not agreeing about the child's profile type or no eligible profile being scored from one parent. A total of 137 comparisons were possible between parents and of these, 21% showed full agreement, 9% partial agreement and 69% non-agreement. These results were consistent with the earlier low agreement shown between parents when scores on

individual scales were used to classify children as cases or non-cases. It appears that as well as showing limited agreement about whether or not a child is a "case", parents also show limited agreement about the broader nature of their child's problems.

The number of profiles identified by individual parents is shown in the horizontal and vertical columns in Tables 5.15 to 5.18. Only profiles identified from a parent checklist which met the dual criteria of having an intraclass correlation of $\geq$0.35 with a standard profile, and a Total Behaviour Problem score above the recommended cutoff, are identified in these columns. The Partial Agreements and some Non-Agreements within each table included profiles where the Total Behavior Problem Score did not exceed the recommended cutoff score. As a result, in some tables the number of Full Agreements, Partial Agreements, and Non-agreements within the body of the table do not equal the number of profiles in the horizontal and vertical columns.

Although the number of profiles was small it can be seen that in three of the four age and sex groups the most frequently identified profile differed for mothers and fathers. For example, for the younger male group, the profile most frequently identified by fathers was the Schizoid Profile. For mothers, the most frequently identified profile was the Somatic Complaints Profile. These results suggest that if information about children's behaviour is collected from only one parent, some disorders may be under-reported.

5.3 COMPARISON OF ITEM SCORES REPORTED BY DIFFERENT INFORMANTS

The final section in this chapter reports the level of agreement between the scores on individual behaviour problem items reported by mothers, fathers, children and teachers. In addition, the percentage of these informants who endorsed individual items is reported.

5.31 Subjects

The subjects for this comparison were the 336 children for whom completed questionnaires were available from two parents, a teacher and the child. Use of this sample ensured that comparisons between reports from the different informants were always based on the same group of children.

5.32 Data Analysis

The approach used to analyze results in this section is modelled on that employed by Herjanic and Reich (1982), and Costello et al. (1984). It should be noted, however, that these earlier studies were restricted to comparisons of reports obtained from mothers and children who had been referred to psychiatric or paediatric clinics. It is plausible that the level of agreement between mothers and children attending clinics, and the frequency with which behaviour problems are reported by such mothers and children, may well be different from that which would be found in the general population. In addition, analyses in this section extend earlier work by reporting not only on results obtained from mothers and children but also results from fathers and teachers.

Comparisons are reported for scores on the 89 items present on the CBCL, TRF, and YSR. Prior to the comparisons, the responses "somewhat or sometimes true" and "very true" were collapsed to form a single category. This preliminary step was necessary because on approximately 50% of items there were insufficient scores in the "very true" category to allow for statistical comparisons of scores using a 3x3 contingency table. Once again, the kappa statistic was used to identify the level of agreement between scores, and McNemar's statistic was used to identify asymmetries. The definitions for high, moderate, low, and non-agreement are the same as those identified by Costello et al. (1984) which were used with earlier analyses (0.60 or higher is defined as "high", 0.40 to 0.59 as "moderate", and 0.20 to 0.39 as "low" agreement. Kappas below 0.20 are considered to reflect non-agreement).

For the purpose of these comparisons, items described as "Externalizing" or "Internalizing" were items which fell into these categories on both the checklists from which informant responses were being compared.

5.33 Results

Three different approaches were used to investigate the level of agreement between informants with regard to individual items.

In the first approach, the total number of high, moderate, low and non-agreements were identified for all the items, and for items defined as externalizing or internalizing (Tables 5.19 & 5.20). It can be seen that in each case a higher proportion of the comparisons between individual item scores from mothers and fathers achieved high or moderate levels of

agreement than occurred with comparisons between mothers and children, or fathers and children. There were also significantly more high and moderate agreements between mothers and fathers for the externalizing items than for the internalizing items ($X^2=14.4$, df=2, $p<0.001$). A similar pattern was evident for comparisons involving mothers and children, and fathers and children, however, these results were not statistically significant ($p<0.05$).

The second approach compared the items where there was good agreement between informants with the items where there was poor agreement. For each pair of informants, the six items which showed the highest level of agreement and the six items which showed the lowest level of agreement were identified for each age and sex group of children. The two sets of items are shown in tables 5.21 to 5.24 (The full set of comparisons for all items is shown in the Appendix). As the focus of these comparisons was on agreement about the presence of emotional and behavioural problems, the items which focused on physical problems or speech problems were not included in this analysis. In general, it can be seen that items which achieved the highest levels of agreement were those describing behaviours which were obvious and could be seen by all the informants. They were also items which could be easily understood and required only limited judgements about their severity or the time over which they had occurred. This pattern was evident not only for mothers and children but also for other pairs of informants. Thus, items such as "Poor school work", "Bites fingernails" and "Poorly co-ordinated" consistently achieved high levels of agreement for all pairs of informants.

The direction of the asymmetries between the children and the other informants reported in the Appendix are summarized in Table 5.25. It can be seen that for both externalizing and internalizing items, the pattern was overwhelmingly one of children endorsing items more frequently than the other informants. The six items in each study group which showed the highest levels of asymmetry are reported in Table 5.26. The items primarily described feelings, thoughts and worries of a private nature which might not be known to anyone but the child. They included items such as "Thinks about sex too much" with the older children, and "Doesn't feel guilty", "Too concerned about neatness" and "Worrying" with the younger children. Interestingly, they also included items describing behaviour such as "Argues", "Talks too much", and "Swearing". These are behaviours which parents often feel their adolescent children fail to recognize and correct. The results of this study suggest that children are aware of these behaviours to a greater extent than parents realize.

The final approach focused on the percentage of items in each age and sex group endorsed by parents, children, and teachers. The results are presented in Figure 1 as a series of graphs which make it possible to easily visualize the percentage of items endorsed by the different informants. It should be noted that results are presented only for the items found on the CBCL, YSR and TRF. As a result, some items are not included and this is reflected in the missing item numbers evident in the figure.

Overall, the prevalence of the individual behaviour problems (and physical problems in the case of items 56a to 56g) varied considerably. For example, "Stealing outside the home" (Item 82) was rarely endorsed by any informant while "Argues a lot" (Item 3) was endorsed by over 60% of

parents and children (Figure 1). However, despite the variation in prevalences across the items, within each item there was a very consistent pattern evident with regard to the prevalences reported by parents, children and teachers. For the great majority of items, children reported the highest prevalence of problems, parents the next highest prevalence, and teachers reported the lowest prevalence of problems. This pattern is consistent with the pattern identified in Chapter 4 when the total problem score, externalizing score and internalizing score were used to compare the number of problems reported by different informants.

5.4 DISCUSSION

In this chapter three different categorical approaches were used to study agreement between informants about the presence of childhood problems. Firstly, the level of agreement between informants was reported with regard to case identification, using the cutoff scores on the child behaviour checklists to identify children as cases or non-cases. Secondly, the level of agreement about the children's behaviour profiles identified by parents was described. Thirdly, the individual behaviour items for which there was high and low agreement between informants, and the frequency of endorsements of all the items by different informants were reported.

When comparisons between different informants were undertaken using the approach of dividing children into cases and non-cases, a number of important findings were apparent. Firstly, the number of children identified as cases on the basis of their scores on the Total Behaviour

Problem Scale, Externalizing Scale, or Internalizing Scale, depended on which informant reports were used to evaluate the children's problems. For example, if the traditional approach of using total problem scores reported by mothers and teachers was employed, 23% of the children who would have been identified as cases if all the informants were used, would not have been identified (Table 5.4). Secondly, using the same scores, approximately 50% of the children were identified by only one informant, and no other, as cases. For example, mothers were the only informant who identified 20% of the children who were identified as cases by at least one informant (Table 5.12). Fathers were the only informant who identified 14% of the cases. What needs to be emphasized is the extent to which the parents were identifying two different groups of children as cases. In fact, when identifying cases, the degree of overlap between all the informants was no more than might be expected by chance. A similar pattern was evident when cutoff scores on either the Externalizing or Internalizing Scales were employed. This suggests that parents show poor agreement with regard to case identification, regardless of the type of childhood problem they are being asked to identify.

The lack of overlap between children identified as cases by different informants is a surprising result given that the reports obtained from the different informants were describing the same child and would not have been expected to be independent of one another. Interestingly, however, the result is consistent with the studies reported by Graham (1967), and Mitchell and Shepherd (1966) which focused on the amount of overlap in reports obtained from teachers and parents. As well, the results are consistent with the results from the study by MacMillan et al. (1980) which focused on agreement between teachers and children. In

all these studies there was little overlap between the children identified as cases by parents and teachers. Finally, the result is consistent with the reports by Lobovits and Handal (1985), and Hulbert et al. (1986) which focused on agreement between children and parents. The results reinforce the point made by Hulbert et al. (1986) that high correlations between informant scores are not necessarily associated with high levels of agreement with regard to case identification.

One difference between the results found in this study and the results reported by Lobovits and Handal (1985) is that the latter found a higher prevalence of problems when prevalences were based on information obtained from children. There are two possible reasons for the different results. Firstly, the study by Lobovits and Handel (1985) focused on a relatively young group of clinic-referred children. As mentioned, it is possible that results in such a population may differ from those found with older children living in the general community. Secondly, the assessment methods used in the two studies were different. In the study by Lobovits and Handel (1985), assessment was by means of diagnoses made by clinicians listening to audiotaped interviews of mothers and children. In this study, case identification used cutoff scores on child behaviour checklists.

The results reported in this chapter extend the work of previous investigators by describing the extent of overlap between mothers, fathers, children, and teachers reporting on problems in the same group of children. As highlighted by Jensen, Traylor, Xenakis, and Davis (1988), most epidemiological studies have assumed that either parent can be used interchangeably to report on children's behaviour. Results in this chapter suggest that this assumption is not correct. Instead, the

results suggest that if mothers are used as the informant in epidemiological studies, different children may be identified from when fathers are used to identify children with problems. This is important because many epidemiological studies employ a two-stage procedure in order to identify children with problems (Bird et al., 1988). In the first stage, questionnaires or checklists are used as a cheap method of screening large numbers of children. In the second stage, children identified on the checklists as having problems are assessed in greater detail using more expensive interview-based techniques. It is possible that the children identified by mothers on the checklists may have different characteristics than the children identified by fathers. If so, studies which rely solely on reports from mothers will not provide a complete picture of the problems suffered by all children.

At the present time it is not clear what the implications are if only one parent identifies a child as having problems, as compared to a situation where both parents report that their child has problems (Edelbrock et al., 1986). Several researchers (Schachar, Rutter, & Smith, 1981; McGee, Silva, & Williams, 1984) have highlighted this issue with regard to reports obtained from mothers and teachers. For example, Schachar et al. (1981) studied children reported by mothers and/or teachers as being hyperactive and suggested that children identified by both mothers and teachers have a qualitatively different disorder than those identified by only one of these informants. To date, there appears to be little information concerning the implications for the aetiology, treatment response and prognosis of childhood emotional and behavioural problems when problems are identified by other combinations of informants. For example, is a child's prognosis and treatment response different if only one parent, as compared to both parents, identifies the problems?

Alternatively, what are the implications if only the child reports the problems in the absence of any parent or teacher concerns? Does this situation indicate a different prognosis and treatment response from one in which all informants identify the problems? This is an important issue in view of the limited overlap which appears to exist between children identified as cases by different informants.

The results from comparisons which focused on individual items revealed a similar picture to that reported for mothers and children in earlier studies (Herjanic & Reich, 1982). However, this study extends the work of Herjanic and Reich by including information about agreement from mothers, fathers, children, and teachers. The results showed that there were significantly more high level agreements between mothers and fathers than between either of the parents and their children. This pattern was evident for both externalizing and internalizing problems. For parents and children, the proportion of high, moderate, low, and non agreements was similar for the externalizing and internalizing problems. This was a different pattern from that identified by Herjanic and Reich (1982) who reported higher levels of agreement between parents and children for externalizing problems than internalizing problems. Once again, a possible explanation for the different results is that the two studies focused on different populations of children. It is possible that in clinic-referred families, unlike families in the community, a higher level of agreement exists about the presence of externalizing symptoms than internalizing symptoms and it is this agreement which contributes to a decision that a child should be referred for help. In comparison, lesser agreement may exist with regard to internalizing symptoms because such symptoms are less obvious to external observers such as parents or teachers.

Consistent with reports from other investigators (Herjanic & Reich, 1982), the items in this study which achieved the highest levels of agreement were items which described behaviours that are relatively unambiguous and easily observed. These included "Poor school work", "Poorly co-ordinated" and "Bites fingernails". Items which showed the lowest levels of agreement tended to be those which would require more judgement by informants as to whether or not they should be endorsed. This is consistent with the suggestion by Kashani et al. (1985) that problems which require a judgement by informants as to whether or not they should be reported, may show poorer agreement than less ambiguous problems. Items which described behaviours or thoughts which are less obvious to external observers or items which might have been misunderstood by some informants also showed poorer agreement. Items in the former category included "Feels too guilty" or "Withdrawn" while the latter category included items such as "Obsessions", "Compulsions" and "Strange behaviour". All the items were scored strictly according to the instructions provided by Achenbach and Edelbrock (Achenbach & Edelbrock, 1983; Achenbach & Edelbrock, 1986; Achenbach & Edelbrock, 1987). However, it was apparent from comments added to some of the latter items that children, in particular, were uncertain about the nature of the thoughts or behaviour described by the item. This uncertainty may have contributed to the lower agreement shown between children and the other informants for the latter items. Finally, for the older children, poor agreement was evident between children and the other informants for items describing behaviours which might not be known to others such as "Truancy" and "Alcohol, drugs". Interestingly, there was considerable consistency as to the items which showed poor agreement across different pairs of informants in the different groups. This suggests that earlier

findings reported for mothers and clinic-referred children, may also apply to parents, children and teachers when describing the problems of children living in the community.

Finally, the prevalence of individual problems reported by different informants revealed a pattern consistent with the results reported in Chapter 4. In general, for individual items the highest prevalence was reported by children, parents reported the next highest prevalence and teachers reported the lowest prevalence. What is evident in the results presented in this chapter is the consistency of this pattern across a wide range of different emotional, behavioural and physical problems. Possible reasons for this pattern have already been discussed in Chapter 4. However, the consistency with which children reported higher prevalences in a such a wide range of problem areas reinforces the view that, as discussed by Kashani et al. (1985), children may have a different threshold for when they judge that it is warranted to report a problem as present.

In summary, the results in this chapter suggest that there is little overlap between the children identified as "cases" by different informants using cutoff scores on the child behaviour checklists. This result has important implications for epidemiological studies which have traditionally relied on reports from mothers and children to identify children with problems. As suggested by Edelbrock et al. (1986), more information is needed about the implications for the aetiology, prognosis, and treatment of childhood emotional and behavioural problems when the problems are identified by different informants. Finally, the results suggest that the higher levels of agreement about externalizing

symptoms than internalizing problems found in clinic-referred populations may not be so evident with children living in the community.

CHAPTER 6. INFLUENCE OF THE DEMOGRAPHIC CHARACTERISTICS OF PARENTS, PARENT PSYCHOPATHOLOGY, AND FAMILY PSYCHOLOGICAL ADJUSTMENT ON THE LEVEL OF AGREEMENT BETWEEN PARENTS AND CHILDREN

6.0 INTRODUCTION

Clinicians often assume that parent psychopathology or family dysfunction may contribute to differences in reports about children's behaviour obtained from different family members. However, there have been few studies which have formally investigated the influence of parent psychopathology on the level of agreement between parents and children, and none which have investigated the impact of family dysfunction. As well, those studies which have investigated the influence of parent psychopathology have produced contradictory results. For example, while Moretti et al. (1985) reported that parent psychopathology may bias parent observations of their children, Weissman et al. (1987) have reported better agreement when a parent informant was depressed at the time of their report.

An important limitation of previous studies is that the studies generally obtained reports from only one parent and a child. As a result, it was difficult to identify clearly the relationship between factors such as the presence of self-reported parent psychopathology and parents' reports of their children's behaviour. In addition, most recent studies have included only children who were patients at mental health clinics

(Jensen, Traylor, Xenakis, & Davis, 1988). As discussed in Chapter 1, there are often substantial differences between the social and intrapersonal attributes of people with the same symptoms and disorders in clinic and community populations (Henderson et al., 1981).

A unique feature of this study is that it included a large representative sample of male and female children from the general community and separate reports were obtained from mothers and fathers, as well as from the children. The availability of reports from two parents, both describing the same child, made it possible to compare reports from parents where one parent reported that he or she had psychological problems while the other parent did not report such problems. In addition, it was possible to study separately the relationship between factors "external" to children's behaviour problems, such as parent occupation or education, and reports from mothers and fathers describing the behaviour of both male and female children.

This chapter describes the apparent influence of three broad groups of factors on the level of agreement between parents and children. The factors include demographic characteristics of the parents, parent psychopathology, and psychological adjustment of the family. These particular factors were chosen on the basis of experience in clinical practice where parent psychopathology and family dysfunction, in particular, are thought to be contributors to discrepancies between reports provided by parents and children.

6.1 INFLUENCE OF THE DEMOGRAPHIC CHARACTERISTICS OF PARENTS

The analyses in this section describe the influence of parent's occupation, level of education, and country of birth on the level of agreement between parents and children.

6.11 Data Analysis

In each analysis, parents and their children were assigned to different groups on the basis of the demographic characteristics of the parent being studied. For these analyses it was necessary to combine the four age and sex groups of children in order to ensure that there were sufficient numbers of parents and children in each group. Multiple regression analyses (SAS Institute Inc., 1987) were used to compare the differences between parent and child scores across the various demographic groups. As the differences had been shown to vary with children of different age and sex (see Chapter 4), age and sex were treated as covariates in these analyses. As mentioned in Chapter 4, the differences between parent and child scores were approximately normally distributed.

In the tables, the mean difference between the parent and child problem scores is shown together with the mean scores reported by the parents and children. This makes it possible to identify the contribution of the individual parent and child scores to variations in the mean difference between their scores across the demographic groups. For example, in a number of instances, the scores from parents and children showed parallel

variations across the groups. When this happened, although the scores from both informants differed across the groups, the mean difference between their scores remained the same. On other occasions, only scores from one of the informants differed across the groups and as a consequence the mean difference between the informants varied from group to group. The difference between the parent and child scores was always calculated by subtracting the parent score from the child score.

It should be noted that the scores reported for parents and children in this chapter were based on all the items which normally comprise each scale (i.e. no items were omitted as was done for the purpose of the analyses reported in Chapter 4). This approach was adopted so that the scores in this chapter can be compared with those reported in other studies which have employed the child behaviour checklists. As there were more items included from the CBCL in this chapter, the mean scores from the parents tended to be somewhat higher. As a result, the differences between the scores reported by parents and children are somewhat smaller than the differences reported in Chapter 4.

6.12 Results

6.121 Parent Occupation

Daniel's occupational status scale (Daniel, 1983) was used to assign parents to different groups according to their occupation at the time of the study.

As mentioned earlier, "housewife" is not ranked on the Daniel's scale and all the mothers who listed this as their occupation were assigned to one

group. The remaining mothers were assigned to one of two groups. One group contained mothers whose occupation on the Daniel's scale was ranked in the range 1.0 to 4.9 and the other contained mothers whose occupation was ranked from 5.0 to 6.9. Fathers were assigned to one of three groups containing occupations ranked from 1 to 2.9, 3 to 4.9, and 5.0 to 6.9 on the Daniel's scale. As noted, 1 to 2.9 equates to occupations such as medical specialist, barrister and bank manager, 3 to 4.9 includes occupations such as computer programmer, and electrician, and 5.0 to 6.9 includes occupations such as typist, junior clerk and barman.

a) Mothers: The mean difference between the total behaviour problem scores reported by mothers and children did not vary greatly across the groups, defined on the basis of maternal occupation (Table 6.1). It can be seen that the largest difference occurred in the group containing mothers whose occupations were rated as more prestigious on the Daniel's scale, and the smallest difference was in the group containing mothers whose occupations were rated as less prestigious. The mean difference in the group containing mothers who listed their occupation as "housewife" was intermediate between the other two groups. The major contribution to the variation in the size of the difference between mother and child scores across the groups came from the scores reported by the mothers. As opposed to the difference between the mother and child scores, the total behaviour problem scores reported specifically by the mothers varied significantly ($p<0.02$) across the occupational groups. In contrast, the scores reported by the children did not vary significantly across the groups.

The externalizing and internalizing scores reported by the mothers and children are also shown in Table 6.1. The variation in the difference

between the mother and child externalizing scores across the groups was relatively small. However, the externalizing scores reported specifically by the mothers, but not those reported by the children, varied significantly across the groups. On the Internalizing Scale, the difference between mother and child scores was greatest in the group containing mothers with more prestigious occupations with little variation between the other two groups.

A limitation of the Daniel's scale is that no rating of prestige is provided for the occupation of "housewife". This occupation was endorsed by 38% of the mothers in the study. In order to address the problem this presented, the comparison of differences between mother and child scores across occupational groups was repeated using the occupation of fathers as the basis for classifying all mothers and children into groups of different occupational status (the groups were defined in the same way as those shown for the fathers in Table 6.2). The mean difference between the mother and child scores, and the individual scores obtained from mothers and children were then compared across these groups. The results of these analyses were consistent with the comparisons based on the mother's occupation. The difference between the mother and child scores was greatest in families in which the father occupied the most prestigious occupation. As well, the total behaviour problem score and the externalizing score reported by the mothers, but not those reported by the children, showed significant variations ($p<0.05$) across the groups. The highest scores were reported by mothers in families where the father was employed the least prestigious occupations. The lowest scores were reported by mothers in families where the father was employed in the most prestigious occupations.

b) Fathers: The size of the difference between the father and child scores did not vary significantly across the three groups on any of the three behaviour problem scales (Table 6.2). This occurred because although the problem scores reported by both the fathers and children were consistently higher in the group containing the fathers with the least prestigious occupations, the variations across the groups for both informants were similar. As a result, the size of the difference between their scores was similar in each group.

In summary, the discrepancies between the scores reported by the parents and children did not show large differences across the occupational groups. In part, this was because the individual scores from both the parents and children varied in a parallel fashion across the groups. As a result, the discrepancies between the parent and child scores in the three occupational groups was similar. However, in some instances the number of problems reported by the mothers, but not those reported by the children, differed significantly across the groups.

6.122 Parent Education

Although parent occupation and education are at times combined to produce a single index of "social class" or "occupational status", for the purpose of this study, analyses involving parent education and occupation are presented separately. This was done because the estimated Pearson correlations between education and occupation were only r=0.4 for mothers and r=0.5 for fathers.

a) Mothers: The mean difference between the problem scores reported by mothers and children in the groups based on the level of maternal education are shown in Table 6.3. Only the discrepancy between the mother and child externalizing scores differed significantly across the three groups. However, the total problem score, externalizing score, and the internalizing score reported by mothers varied significantly across the groups. In contrast, none of the scores reported by the children varied significantly across the groups.

b) Fathers: The mean difference between father and child scores in the groups based on the level of education of the father are reported in Table 6.4. There was little variation across the groups in the size of the mean difference between father and child scores on any of the behaviour scales. This was due to the fact that both father and child scores varied in a parallel fashion across the groups.

In summary, the mean difference between parent and child scores did not vary greatly between the groups containing parents who had completed different levels of formal education. This suggests that parent education is not a major factor influencing the difference between reports of children's behaviour obtained from parents and children. In contrast, level of education was associated with significant differences in the number of problems reported specifically by mothers, but not those reported by the children. This is a similar pattern to that identified earlier in the results which described the effect of maternal occupation on problem scores reported by mothers and children.

6.123 Country of Birth

For these comparisons, parents and their children were assigned to one of three groups according to the country of birth of the parent. This was done separately for the mothers and the fathers (Tables 6.5 & 6.6).

The mean differences between parent and child scores did not vary greatly across the three groups and there was no consistent direction to the pattern of variations (Tables 6.5 & 6.6). Only the mean difference between the mother and child scores on the Externalizing Scale varied significantly across the groups.

6.2 INFLUENCE OF PARENT PSYCHOPATHOLOGY AND FAMILY ADJUSTMENT

This section describes the influence of parent psychopathology and family adjustment on the parents' reports of their children's behaviour. In addition, the section describes the influence of parent psychopathology and family adjustment on the level of agreement between the problem scores reported by parents and children.

6.21 Procedure

The analyses reported in this section focus on the total problem scores obtained from the Child Behavior Checklists completed by the mothers (the mother-reported problem score) or the fathers (the father-reported problem score), and the total problem score on the Youth Self-Report completed by the children (the child-reported problem score).

The psychological adjustment of parents was assessed by means of the General Health Questionnaire (GHQ) (12 item version). The four point response scale from this questionnaire can be scored in two ways (Goldberg, 1978, p. 8). It can be treated as a multiple-response scale with items scored from 1 to 4 or it can be treated as a bimodal response scale with items scored as 0 or 1. Goldberg (1978, p. 8) suggests that the advantage of the latter scoring procedure is that it "eliminates any errors due to 'end-users' and 'middle-users'". When the bimodal scoring procedure is used with the GHQ, recommended cut-off scores are available which can be used to identify subjects with psychological problems. For the purpose of the categorical analyses reported in this section, parents who scored above the cutoff score on the GHQ were classified as "disturbed" and those who scored below the cutoff were classified as "healthy".

Family psychological adjustment was assessed by means of the General Functioning Scale of the Family Assessement Device (FAD) (Epstein et al., 1983). This questionnaire was completed by all parents and the 14-15 year old children. Unlike the GHQ, there are no established cutoff scores which can be used with the General Functioning Scale to categorize families as healthy or disturbed. However, Byles et al. (1988) have suggested that the 90th percentile is a useful cutting point for analyses involving scores on the FAD. When this percentile was used as a cutoff score in the study by Byles et al. (1988), families scoring above the cutoff were significantly more likely to be suffering from problems such as alcohol abuse, marital violence and parent(s) arrest, all of which are thought to be associated with family dysfunction. At present, there are no guidelines available which can be used to combine scores from

different family members into a single rating of family adjustment summarizing the views of all family members. In light of this, the FAD scores from mothers and fathers were analyzed separately in this section.

6.22 Data Analysis

The first set of analyses reported in this section focuses on the influence of parent psychopathology (based on the parent's self-reported GHQ score derived using the multi-response scoring procedure) and the parent's perception of their family's adjustment (based on the parent's FAD score) on the parent's perception of their child's problems (the parent-reported problem score). Initially, simple regression analyses were used to investigate the univariate relationship between the parent-reported problem score, and the parent GHQ score and FAD score. Subsequently, multiple regression analyses were used to test for the effect of parent psychopathology on the parent-reported problem score, allowing for the influence of family adjustment.

The second set of analyses reported in this section focuses on the relationship between the parent-reported problem scores and the child-reported problem scores. In these analyses, regression analyses were used to study the relationship between the problem scores reported by children and their parents, allowing for the effect of parent psychopathology and family adjustment. Initially, simple regression analyses were used to study the univariate relationships between the child-reported problem score, the parent-reported problem score, the parent's GHQ score and the parent's FAD score. Subsequently, multiple regression analyses were used to test for the significance of the association between the child-reported problem score and the parent-

reported problem score, allowing for the influence of parent psychopathology and family adjustment. In these analyses the dependent variable was the child-reported problem score, and the independent variables were the parent-reported problem score, the parent's GHQ score, and the parent's FAD score.

The advantage of using regression analyses is that they allow the original behaviour scores to be retained and there is no loss of information as a consequence of collapsing categories. They also provide information about both the size of the correlation between the dependent and independent variables, and the size of the unit change of the variables included in the analyses. A disadvantage of regression analyses is that it can be difficult for clinicians, not trained in the use of multivariate statistics, to "visualize" the meaning of the results reported using this approach. As well, clinicians tend to employ a more categorical approach to the classification of adult psychopathology and family adjustment. Finally, in a more statistical context, the regression approach requires more assumptions about the distribution of scores and homogeneity of variance.

In order to make the results reported in this chapter more relevant to clinicians, categorical analyses have been included in the presentation of results. In the categorical approach, parents or families were assigned to one of two groups labelled "healthy" or "disturbed", on the basis of parent scores on the GHQ or FAD. The categorical analyses then compared the difference between the parent and child scores in the healthy and disturbed groups. As described earlier, in some instances, the scores from parents and children showed parallel variations across the healthy and disturbed groups. As a result, although the mean scores

from both informants varied in the two groups, the mean difference between scores in the two groups remained constant. On other occasions, only the mean scores from one informant varied across the groups and thus the mean difference between the informants varied from group to group.

In the presentation of results of the categorical analyses, use is made of simple graphs to highlight the mean scores reported by parents and children in the healthy and disturbed groups. This approach was chosen because it enables the reader to see easily the influence of a score from one informant on the mean difference between scores from two informants. The Kolmogorov-Smirnov statistic was used to test for the significance of differences between scores obtained from parents and children in the disturbed and healthy groups. This statistic was chosen because, as mentioned, the distributions of behaviour scores were positively skewed. The differences between parent and child scores showed an approximately normal distribution and t-tests were used to test for the significance of differences in the mean difference between parent and child scores.

6.23 Results

6.231 Parent-reported Problem Scores

The first set of results reported in this section focuses on the effect of parent's perception of their own psychological adjustment (GHQ score), and that of their family (FAD score), on the parent's perception of their child's behaviour (the parent-reported problem score). In these analyses the parent-reported problem score is the dependent variable and the parent's GHQ and FAD scores are the independent variables.

a) Mothers: The results of simple regression analyses testing for the effect of GHQ and FAD scores (obtained from the mother) on the mother-reported problem score are shown in Table 6.7. It can be seen that, except in the group containing the older female children, the GHQ score had a significant positive effect on the mother-reported problem score. Interestingly, the effect of the GHQ score was consistently larger for the scores describing the male children than for scores describing the female children. In addition, the proportion of variance accounted for by the GHQ score was larger in the groups containing the male children. The FAD score had a significant effect on the mother-reported problem score in all the groups except that containing the older male children. It should be noted that the relatively high regression coefficients for the FAD scores reflected the fact that these scores occupied a somewhat narrow range. For all the groups, however, both the mother's GHQ and FAD scores explain only a relatively small proportion of the variance (r^2) in the mother-reported problem score.

The results of the multiple regression analyses testing for the effect of the mother's GHQ score on the mother-reported problem score, while allowing for the effect of the FAD score, are shown in Table 6.8. It can be seen that when allowance is made for the effect of the FAD score, the size of the effect of the GHQ score on the mother-reported problem score is somewhat less than that found when only the univariate relationship between GHQ score and mother-reported score is tested. However, both the size of the effect, and the proportion of variance in the mother-reported problem score accounted for by the GHQ score is still greater in the groups containing the male children than in the groups containing the female children.

b) Fathers: The results of the simple regression analyses testing for the effect of father's GHQ and FAD scores on the father-reported problem score are shown in Table 6.9. The results of these analyses were similar to those described for mothers. Once again, except in the group containing the older female children, the GHQ score had a significant positive effect on the father-reported problem score. Although the effect of the GHQ score was larger for reports describing the older male children than the older female children, this difference was less evident with the reports describing the younger male and female children. A similar pattern can be seen with the proportion of variance accounted for by the GHQ score. A further difference between the results obtained with mothers and fathers occurred in the group containing the older male children. In this group, the FAD score reported by fathers had a significant positive effect on the father-reported problem score. In addition, the FAD score accounted for 29% of the variance in the father-reported problem score. No significant effects were found with the equivalent mother scores.

The results from the multiple regression analyses testing for the effect of the father's GHQ score on the father-reported problem score, while allowing for the effect of the father's FAD score, are shown in Table 6.10. The differences between the results for the male and female groups showed a less consistent pattern than was found with the mother scores. Once again, however, with the exception of the FAD score in the older male group, the proportion of variance in the father-reported problem score explained by the father's GHQ and FAD scores, was relatively small ($\leq$0.14).

An important strength of this study is the availability of reports describing the same child obtained from different parents in the same family. This made it possible to compare the scores obtained from parents, both describing the same child, where one parent rated him or herself in the disturbed range on the GHQ, and the other parent rated him or herself in the healthy range. The results of this comparison are reported in Table 6.11. It can be seen that for the male children, the difference between the scores obtained from two parents in the same family was significantly greater when one parent scored him or herself in the disturbed range of the GHQ. This difference occurred because the disturbed parent, regardless of whether they were a father or mother, consistently reported more problems with their child than the healthy parent. A similar pattern was evident with the reports describing the female children. However, for the female children, the difference between the parent scores in families containing a "disturbed" mother did not vary greatly from the difference between parent scores in families where both parents were "healthy" (Table 6.11). This is consistent with the results reported using multiple regression analyses where it was noted that the effect of the mother's GHQ score on the mother-reported problem score was somewhat less for the female children than for the male children.

6.232 Child-reported Problem Scores

The set of results reported in this section focuses on the relationship between the child's perception of their problems (the child-reported problem score) and the parent's perception of their child's problems (the parent-reported problem score), allowing for the effect of parent psychopathology (GHQ score) and family adjustment (FAD score). In this

set of analyses, the child-reported problem score is the dependent variable, and the parent-reported problem score, parent GHQ score, and parent FAD score are the independent variables.

a) Mothers: Initially, simple regression analyses were used to test for the significance of the effect of the mother-reported problem score, GHQ score, and the FAD score, on the child-reported problem score. Results from these analyses are reported in Table 6.12. It can be seen that the mother-reported problem score had a significant positive effect on the child-reported problem score in each of the four age and sex groups of children. In addition, the size of the effect was similar in the four groups. The mother's GHQ score had a significant effect on the child-reported problem score for the male children but not for the female children. Finally, the mother's FAD score had a significant effect of on the child-reported problem score in all four age and sex groups.

The proportion of variance in the child-reported problem score (r^2) which could be accounted for by variance in the mother-reported problem score, GHQ score and FAD score are shown in Table 6.12. It can be seen the proportion of variance in the child-reported problem score accounted for by the mother-reported problem score was 45% for the younger male children but only 17% for the older male children. A different pattern was evident for the female children. For these children, the proportion of variance in the child-reported problem score accounted for by the mother-reported score in the younger group was 28%, while 35% was accounted for in the older group. The proportion of variance in the child-reported score accounted for by the mother's GHQ and FAD scores was generally small in all the groups.

Multiple regression analyses were used to identify the effect of the mother-reported problem score on the child-reported problem score, allowing for the effect of the mother's GHQ score and FAD score. The results of these analyses are shown in Table 6.13. Two findings are evident. Firstly, it can be seen that the size of the effect of the mother-reported problem score on the child-reported score in the multiple regression analyses was similar to that found using the simple regression analyses. This result is consistent with the generally low r^2 ($\leq$0.15) shown in Table 6.7 which reported the effect of the mother's GHQ score and FAD score on the mother-reported problem score. Secondly, the inclusion of the mother's FAD and GHQ scores in the regression model added little to the proportion of the total variance in the child-reported problem score which could be accounted for, beyond that already accounted for by the mother-reported score. For the total behaviour problem score in the group containing the 10-11 year old male children, the slight decline in the r^2 (0.45 to 0.40) from that found with the simple regression was due to the loss of subjects in the multiple regression analysis.

b) Fathers: The results of the simple regression analyses testing for the effect of the father-reported problem score, GHQ score, and FAD score on the child-reported problems score are shown in Table 6.14. The results were similar to those found with the equivalent mother scores. In particular, the effect of the father-reported problem score was significant in all four groups and was of a similar magnitude to that of the mother-reported score. Interestingly, the father's GHQ score had significant effect on the child-reported problem score for the older children but not for the younger children. Finally, the father's FAD

score had a significant effect on the child-reported problem score in three of the four age and sex groups.

The proportion of variance (r^2) in the child-reported problem score which could be accounted for by the father-reported problem score, GHQ score and FAD score are also shown in Table 6.14. It can be seen that the proportion of variance in the child-reported problem score which was accounted for by variance in the father's score ranged from 42% for the younger male children to 19% for the two groups of older children. The results for the male children are similar to those found with the mothers. However, in the group containing the older female children, the father-reported problem score accounted for only 19% of the variance in the child-reported score. This was considerably smaller than the 35% accounted for by the mother-reported problem score. In general, the proportion of variance in the child-reported problem score accounted for by the father's GHQ and FAD scores was small. The exception to this pattern occurred in the group containing the older female children. In this group, the father's GHQ score accounted for 15% of the variance in the child-reported score.

The results of the multiple regression analyses used to identify the effect of father-reported problem score on the child-reported score, allowing for the effect of the father's GHQ and FAD scores are shown in Table 6.15. In the groups containing the male children, the results of these analyses were similar to those reported for mothers. However, for the older female children, the effect of the father-reported problem score on the child-reported score was considerably smaller than that found with the mothers. As well, in this group, the father-reported

score accounted for only 8% of the variance in the child-reported problem score, compared to the 37% accounted for by the mother-reported score.

Consistent with the results from the mothers, inclusion of the father's FAD and GHQ scores in the regression models generally added little to the proportion of variance which could be accounted for in the child-reported problem score, beyond that which was already accounted for by the father-reported problem score. The one exception to this pattern occurred in the group containing the older female children where 17% of the variance of the child-reported problem score could be accounted for by variance in the father's GHQ score.

6.233 Results of Categorical Analyses

The mean differences between the total behaviour problem scores obtained from mothers and children are shown in Figure 6.1. The lines drawn between the mean scores in the healthy and disturbed groups are included to enable the reader to locate easily scores from the informants in each group. The significance levels from the Kolomogorov-Smirnov statistic and the t-tests, testing for the significance of differences between scores in the two groups are shown in the figures. In general, it can be seen that the size of the mean difference between the mother and child scores was slightly smaller in the disturbed group than in the healthy group. However, the variation between the groups was small and none of the differences between the groups reached a $p<0.05$ level of significance. This reflected the fact that the behaviour scores reported by both mothers and children were consistently higher in the group containing the disturbed mothers (Figure 6.1), although the difference

between the scores reported by the mothers in the two groups was somewhat larger than the difference between the scores reported by the children.

The mean difference between the behaviour problem scores reported by fathers and children in the healthy and disturbed groups are shown in Figure 6.2. The pattern was similar to that found with mothers and children, except for the group containing the 14-15 year old female children. In the latter group, the mean difference between the scores reported by the fathers and children in the disturbed group was significantly greater than that in the healthy group. It can be readily seen that this was primarily due to the influence of the difference in the scores reported by the children in the two groups. This is consistent with the result obtained using the regression analyses.

The mean scores from parents and children in families classified as healthy or disturbed on the basis of the mother's FAD score are shown in Figure 6.3. There was little variation in the mean difference between mother and child scores in the two groups. This reflected the fact that the behaviour scores reported by mothers and children varied in a similar fashion across the groups. In general, the behaviour scores reported by both mothers and children were higher in the group identified by the mother's FAD score as containing the more disturbed families.

Except for the group containing 10-11 year old male children, the mean differences between the scores reported by fathers and children in the healthy and disturbed groups (defined on the basis of the father's FAD score) varied little (Figure 6.4). However, in contrast to the relatively small differences between scores obtained from mothers in healthy and disturbed families, the differences between father scores in

the two groups were all significant at the $p<0.05$ level. Differences between the FAD scores reported by the children were in the same direction as those of the fathers, but somewhat smaller in magnitude.

6.3 DISCUSSION

The results in this chapter described the apparent influence of the demographic characteristics of parents, parent psychopathology, and family psychological adjustment, on the level of agreement between parents and children. In addition, the effect of these factors on the scores obtained from both the parents and the children was reported.

The level of agreement between parents and children varied somewhat across the different demographic groups. Where significant variations did occur they were always for discrepancies between the mother and child scores. None of the discrepancies between the father and child scores showed significant variations across the demographic groups. Although this result was partly due to the smaller number of fathers in the study, the magnitude of the variation across the groups between the mother and child scores was consistently greater than that between the father and child scores (it should be noted that the demographic groups in tables 6.1 and 6.2 are defined differently).

In some groups, demographic characteristics appeared to influence the report from one informant but not from another. For example, mothers with less education reported significantly more problems in their children than were reported by mothers who had achieved higher levels of education. In contrast, there were no significant variations in the

child scores across the three groups defined on the basis of maternal education. This pattern was also evident across the groups defined on the basis of maternal occupational status. Once again, there were significant variations between the mother scores across the three groups but no significant variations between the child scores. Finally, a similar pattern was evident with the reports from parents and children describing the 10-11 year old children of the "healthy" and "disturbed" parents (Figures 6.1 & 6.2). For these children, there was a significant difference between the parent-reported problem scores describing the children in the "healthy" and "disturbed" groups but no significant difference between the child-reported problem scores in the two groups. These results suggest that the apparent relationship between variables such as maternal education, maternal occupation, or parental psychopathology, and childhood behaviour problems, may differ depending on whether the mother or the child is used as the informant to describe the child's behaviour problems. As discussed in Chapter 1, the possibility that some factors "external" to children's behaviour may influence specifically reports about children's emotional and behavioural problems obtained from one informant, but not from others is an important issue. If true, this would suggest that some relationships between external factors, and childhood behaviour problems are "informant specific". That is, the relationships are evident only for reports obtained from a specific informant.

There are several possible explanations for informant specific relationships. Henderson (1981) has suggested that the perceived adequacy of social relationships in the face of adversity, as distinct from a lack of relationships, is important for the prediction of later neurotic symptoms in adults. This finding may have relevance for

informant specific relationships identified in reports of childhood behaviour problems. For example, some children may become unhappy and withdrawn because they perceive that they have few friends, despite the fact that their parents report that the children have adequate social relationships. If this occurred consistently, a different relationship between friendship patterns, and childhood unhappiness and withdrawal might be identified if children were used as informants, as compared to that which would be identified if parents were used as informants. More generally, it is possible that a broader understanding of the aetiology of childhood disorders may emerge if greater use is made of children as informants about their problems and life circumstances, as compared to that which will be found if reliance is placed primarily on reports from parents.

More recently, Rutter (1987) has highlighted the relevance of children's cognition as a mediating factor which may influence children's susceptibility to the effect of different life experiences. However, a child's level of cognitive development may also influence their perception of their emotional or behaviour problems, and events influencing their development. As a result, children may have a different perception of external stressors and the influence of these stressors on their behaviour. If so, reports obtained from children may be important in order to obtain a wider knowledge of the aetiology of childhood disorders.

As well as having important implications for a broader understanding of the aetiology of childhood disorders, the possibility of informant specific relationships also has important implications for the treatment and prevention of childhood emotional and behaviour problems. For

example, if some children are susceptible to becoming depressed or suicidal when they perceive that they have no friends, regardless of their parents' or teachers' observations, this suggests that preventative measures might have to rely more on reports from children in order to identify those children who are at greatest risk for later depression or suicide. Reports of children's friendships obtained from parents or teachers may prove less valuable in such circumstances. Similarly, greater weight may need to be given to reports from children when attempting to assess the effectiveness of treatment programs aimed at providing help for depressed or suicidal children. Finally, where reports from different informants all appear to be influenced in a similar way by a factor which is "external" to children's behaviour, it is possible that this is evidence of a particularly strong relationship between this factor and childhood emotional and behavioural problems. Such factors may warrant more careful attention by clinicians and researchers involved in the assessment of childhood problems.

Several interesting results were apparent in the analyses which focused on the effect of parent psychopathology on parents' reports of their children's behaviour problems. Firstly, the effects were not the same for reports from mothers and fathers. As well, the effects differed for reports describing male and female children. For example, the mother's GHQ scores appeared to have a larger effect on mother reports describing male children than the reports describing female children. This pattern applied with both the older and younger children. The proportion of variance accounted for by the mother's GHQ score was also greater for reports describing male children than reports describing female children. Both these patterns were less evident with the reports obtained from fathers. Although the effects of the GHQ scores were generally small, it

is interesting to note that they were similar to the effects identified with the mother reports describing male and female children in the study by Jensen, Xenakis, Davis, and Degroot (1988).

It is important to note that both this study, and that of Jensen, Xenakis, Davis, and Degroot (1988), employed a cross-sectional design. As a result, it is not possible to establish the direction of the cause and effect relationship between variables showing a significant association. For example, it seems reasonable to suggest that a mother who feels psychologically distressed, will report more behaviour problems in her child. However, it is equally possible that when a mother perceives that her child has many behaviour problems, this causes psychological distress for the mother. There is a great need for prospective studies which could clarify the direction of such cause and effect relationships.

In the past, studies investigating the influence of parent psychopathology on reports of children's behaviour generally obtained reports from only one parent and a child. This has made it difficult to interpret the meaning of results obtained in these studies (Jensen, Traylor, Xenakis, & Davis, 1988). For example, if it was found that parents who reported themselves as having more psychological distress, also reported more behaviour problems in their children, two different interpretations could be made with regard to the meaning of the results. First, it is possible that the results reflected an accurate picture of the relationship between children's behaviour problems and parental psychopathology. That is, if a parent has psychological problems it is likely that their child will have emotional and behavioural problems. Equally, however, it is possible that children with a psychologically

distressed parent do not have more problems than children with two healthy parents. Instead, the presence of psychological distress in the parent may be associated with a perception by the parent, that their child has emotional and behavioural problems. As a result, the report provided by a psychologically distressed parent may differ from that provided about the same child by a parent who is free of psychological problems.

A strength of this study is that it obtained separate reports describing the same child from two parents in the same family. The study also included a large sample of both male and female children living in the community. This made it possible to compare the number of childhood problems reported for male and female children in families where one parent rated him or herself as being "disturbed" and the other rated him or herself as "healthy". The results of these comparisons showed that, particularly for reports describing male children, mothers or fathers who rated themselves as being "disturbed" consistently reported more externalizing and internalizing problems in their children than were reported by their "healthy" spouse. This is a particularly interesting result because the two parents were describing the same child. It suggests that the presence of psychological distress in a parent is associated with a perception by the parent that their child has more emotional and behavioural problems.

There are two possible explanations for the difference between the number of problems reported by the "disturbed" and "healthy" parents. Firstly, as suggested by earlier authors (Moretti et al., 1985), it is plausible that parents who perceive themselves as having more problems may also perceive more problems in their child. This may occur because distressed

parents project some of their distress onto their child. An alternative explanation is that the children of "disturbed" and "healthy" parents behave differently with each parent. For example, it is possible that children are more upset and irritable when in the presence of an unhappy or distressed parent than when they are with their healthy parent. If true, this would provide an alternative explanation for why a distressed parent reports more problems in their child than a healthy parent.

The difference between the reports from "disturbed" and "healthy" parents is an important issue because clinicians are often confronted with situations where not only does a referred child appear to have problems, but so does one or both of his or her parents. In these circumstances it can be difficult for clinicians to know how they should weight reports from the different parents and the child in order to formulate a treatment plan for the child. Although the results presented in this chapter cannot fully explain why the differences exist in the reports from "disturbed" and "healthy" parents, they do provide information which clarifies the extent and direction of the differences. Further work is needed to understand more clearly why the differences between distressed and healthy parents arise and the implication of such differences for a child's prognosis. Finally, the results suggest that studies investigating the relationship between parent psychopathology and childhood behaviour problems should obtain reports from multiple informants. If reports are obtained only from one parent and he or she has psychological problems, it is likely that a higher prevalence of problems in a child will be identified than will occur if information is obtained from other informants. As discussed earlier, it is also possible that the nature of the relationship identified between parent

psychopathology and childhood behaviour problems will differ, depending on which informant is employed to report on these variables.

Despite the rapid growth of family therapies during the last decade, the relationship between family adjustment and childhood behavioural problems has received less attention than the relationship between marital adjustment and childhood problems. In part, this reflects a lack of suitable measures available to assess family adjustment (Sawyer et al., 1988). The results in this chapter suggest that there is a significant relationship between parent perception of their family's adjustment and their perception of behaviour problems in their child. For example, for reports from both parents, the FAD score had a significant positive effect on the parent reports of problems in the younger male and female children. Interestingly, for reports from both parents, in the younger group of children the size of the effect and the proportion of variance accounted for in the parent-reported problem scores was greater for the male children than for the female children. A similar pattern was evident for reports from the fathers with the older children. However, for reports from the mothers, the pattern was reversed with the FAD score having a larger effect on the mother reports describing the female children than the male children.

The results from the analyses which focused on the relationship between the children's reports of their problems and the parent reports of the children's problems, taking into account parent psychopathology and family adjustment, highlighted a number of important issues. Firstly, the results were again different for mothers and fathers when describing male and female children. The amount of variance in the scores from the older male children accounted for by the scores from their parents, was

less than that which applied with the younger male children. A possible reason for the difference is that parents of older male children are less familiar with the problems of their children because the older children spend more time away from the family home. This result has important clinical implications as it suggests that for the assessment of older male children, reports from parents may provide only a limited substitute for reports from the older male children themselves.

The pattern of results obtained with the female children differed from that obtained with the males. In particular, the amount of variance in the child-reported problem score accounted for by the mother-reported problem score was slightly greater for the older female children than for the younger female children. This was the reverse of the pattern found with the reports from the mothers describing male children. Although the results of reports from fathers were similar for the younger male and female children, the fathers' problem scores accounted for a particularly low proportion of the variance in the scores reported by the older female children. It is possible that this finding reflects a lack of knowledge by fathers about the behaviour of their older daughters.

Overall, the extent to which parent reports accurately predict reports from their children varied, even after parental psychopathology and parent perception of their family's adjustment was taken into account. Both Jensen Traylor, Xenakis, and Davis (1988), and Schaughency and Lahey (1985) have drawn attention to the possibility that the sex of the parent, and of the child being described, may influence their reports of children's behaviour problems. The results in this study provide further support for this possibility. The issue is important because, as pointed out by Jensen, Traylor, Xenakis, and Davis (1988), it has been

assumed that either parent can be used as the reporter in studies of childhood psychopathology. The results reported in this chapter do not support this assumption and instead suggest that caution is necessary when interpreting the significance of reports from individual parents, particularly if the parent is suffering from psychological distress.

It is, perhaps, not surprising that parents appeared less familiar with the problems of their older children. Children who are 14-15 years old often spend a considerable amount of time away from the family home and it seems reasonable to assume that their parents may be less familiar with their problems than parents of younger children. Other investigators have also noted that with this older age group many emotional and behaviour problems are revealed only when reports are obtained directly from the adolescents (Rutter, Graham, Chadwick, & Yule, 1976). The different results obtained with mothers and fathers when describing the older females are interesting and have direct clinical relevance. In particular, it appears that mothers may be more familiar with their daughters' problems than are fathers. This finding is not a surprising result but rather one that confirms a view held by many clinicians. It is possible that the result reflects a better level of communication between mothers and older daughters than occurs between the daughters and their fathers.

In summary, the results reported in this chapter highlight the possibility that reports obtained from one informant may be influenced in different ways by factors "external" to a child's behaviour, than reports obtained from other informants. The possibility that the apparent effect of some external factors on children's behaviour are "informant specific" is important. In the past, most studies relied on reports from mothers

and teachers when investigating the possible influence of important external factors, such as poverty or parental psychopathology, on children's behaviour. It had been assumed that the pattern of results found with reports from these informants accurately reflected the results which would be obtained with reports from any other informant. The results from this study suggest that it may not have been correct to make this assumption. Instead, it appears that different relationships may be identified between children's behaviour and other factors, depending on the informant used to describe the children's behaviour.

Finally, the results suggest that the capacity of parents to predict what their children themselves would report about their behaviour problems varies, depending on the age and sex of the child, and the sex of the parent. This finding has important clinical implications because it suggests that there are significant limitations to the comprehensiveness of assessments which are based on reports from only a limited number of informants. As well, it again emphasizes the importance of obtaining reports about their behaviour from children themselves.

CHAPTER 7. A COMPARISON OF REPORTS DESCRIBING CHILDHOOD EMOTIONAL AND BEHAVIOURAL PROBLEMS OBTAINED FROM FAMILIES IN THE GENERAL COMMUNITY AND FROM FAMILIES CONTAINING A CHILD REFERRED TO A MENTAL HEALTH CLINIC

7.0 INTRODUCTION

A possible reason for the contradictory results in earlier studies is that they included children drawn from different populations. For example, previous studies have included children whose parents had a psychiatric illness (Weissman et al., 1980), children attending paediatric clinics (Herjanic & Reich, 1982; Reich et al., 1982), children attending mental health clinics (Edelbrock et al., 1986), and children living in the general community (Jensen, Traylor, Xenakis, & Davis, 1988; Jensen, Xenakis, Davis, & Degroot, 1988). It is plausible that the difference between the study populations may explain, in part, some of the differences between results found in these earlier studies.

This chapter compares reports from parents and children describing behaviour problems in children drawn from two different populations. The first population consisted of children living in the general community. The second consisted of children referred to mental health clinics. The chapter focuses on the extent to which the pattern of discrepancies between reports from the parents and children living in the general community also applied with reports describing children who had been referred to mental health clinics.

7.1 METHOD

7.11 Subjects

As described in Chapter 2, two hundred two-parent families who took part in the mail survey were randomly selected for home visits (Table 2.1). During the visits, the participating children and their parents completed child behaviour checklists (CBCL & YSR). This was done to provide a check on the independence of reports obtained during the mail survey. In addition, the home visits provided the opportunity to conduct interview-based assessments with children. The interviews employed the child (DISC) and parent (DISCP) versions of the Diagnostic Interview Schedule for Children (see Section 2.34 in Chapter 2 for a description of the DISC and DISCP). In order to keep the amount of time spent conducting home visits to manageable proportions, the DISC and DISCP interviews were conducted in only half of the families (also randomly selected) chosen for a home visit.

A total of 163 families agreed to a home visit, representing a response rate of 81.5%. The most common reason given by families who did not agree to a home visit was that they were too busy. In two families, the parents had separated since the mail survey and one parent no longer lived with the participating child. Of the 100 families randomly selected for DISC and DISCP interviews, 81 agreed to be interviewed, representing a response rate of 81%. The only significant difference between the demographic characteristics of the community parents and children who participated in the interviews, and the full group of

subjects who participated in the mail survey, was that a smaller proportion of the interviewed mothers identified their occupation as "housewife" (Likelihood Ratio Chi-square = 9.9, df=4, $p<0.04$). This difference reflected the absence of single mothers in the interviewed sample, the majority of whom had listed their occupation as housewife. All of the home visits with community families took place after the families had participated in the mail survey.

The clinic-referred children were aged 10-16 years and had two parents who had lived with the child for not less than 12 months prior to their assessment in the study. All the clinic-referred children had been referred to one of four participating mental health clinics in Adelaide. The clinics included two out-patient clinics, a day-patient clinic, and an inpatient service. The children in this study represented consecutive referrals to these clinics over a period of 18 months. Children with emotional and behavioural problems are referred to the clinics by a wide range of health and welfare services in the community. All referred children are seen by a psychiatrist, psychologist, or psychiatric social worker and the nature of their assessment and treatment varies, depending on the nature of their problem and the philosophy and training of their clinician. The focus on two-parent families in this study allowed comparisons to be made between the reports obtained from two parents who were describing the same child. At the request of the clinicians working in the clinics, children who were referred for assessment following suspected sexual abuse were not included in the study.

A total of 114 children with two parents were referred to the clinics while the study was in progress and 82 (72%) families agreed to participate in the study (Informed consent was sought from both the

children and parents in all the community and clinic families selected for a home visit). The most common reason (N=15) given by clinic families for not wanting to participate in the study was that their family had already had too much contact with doctors or clinics and they did not want to be involved with additional staff. Clinicians requested that a further eight families not be included. The reasons for these requests varied. For example, in one case a clinician was concerned that a child was suicidal and in another the reason for referral was a recent termination of pregnancy which the family wished to remain confidential. Finally, five families were not included because the parents did not speak English. Overall, the sample of clinic-referred children may be under-representative of children living in non-English speaking families and children whose families have had a large number of contacts with doctors and clinics.

The age and sex of participating children in both the community and clinic-referred families are shown in Tables 7.1 and 7.2. Table 7.1 shows the community children who completed only a Youth Self Report. The parents of this group of children completed only a Child Behavior Checklist. Table 7.2 shows the community children who completed a Youth Self-Report and a DISC interview. The parents of this group of children completed both a Child Behavior Checklist and a DISCP interview. The occupation and education level of the parents of the two groups of community children, and of the clinic-referred children are shown in Tables 7.3 and 7.4. It can be seen that there were some differences between the demographic characteristics of the clinic and community families. In particular, the parents in clinic-referred families had less formal education, and a greater proportion of the parents in clinic-referred families had occupations which were defined as less prestigious.

As reported in Chapter 6, however, these factors do not appear to have a large influence on the size of the difference between problems scores reported by parents and children.

7.12 Procedure

The community families selected for home visits were initially contacted by letter and this was followed by a telephone call during which consent for a home visit was sought. A small number of families did not have a telephone and it was necessary to visit the family home to seek their consent to participate in this part of the study.

All the clinic-referred children who met the entry criteria for the study, and their families, were advised about the study at the time of their first clinic appointment. Following their appointment, families were contacted by telephone and their informed consent to participate in the study was sought. For most families, the home visit was completed prior to their second clinic appointment. Although it was not possible to randomize completely the order in which the checklists and interviews were administered, this was done as far as was practicable.

Prior to commencing the home interviews, two research assistants were trained in the use of the DISC and DISCP. A pilot study of 20 parents and children was undertaken to test the inter-rater reliability of the two assistants. The results of this study are reported in the Appendix. It can be seen that the inter-rater reliability in the study was very high. Soon after the home interviews commenced, one of the assistants left the project and as a result, all the interviews were conducted by one research assistant.

7.13 Measures

All the participating children and their parents completed child behaviour checklists (YSR and CBCL) during a home visit. Except where explicitly stated, the standard items which comprise each scale were employed in this chapter (i.e. items on the scales were not omitted as was done in Chapter 4 to ensure that the scales were comprised of the same number of items). All the clinic-referred children and half the children in the community, and their parents, completed DISC and DISCP interviews. Background information about both measures was provided in Chapter 2.

7.14 Data Analysis

The approach to data analysis used in this chapter was the same as that employed in earlier chapters. Specifically, a non-parametric statistic (the Wilcoxon Signed-Ranks Test) was used to test the significance of differences between scores obtained from parents and children. This statistic was chosen because these scores were not normally distributed. The significance of the difference in the discrepancies between parent and child scores across the community and clinic groups, which showed an approximately normal distribution, was tested using t-tests.

7.2 RESULTS

a) Child Behaviour Checklists: The parent and child scores describing children in the clinic and community groups are shown in Tables 7.5 and

7.6. In each case the mean difference reported in the table is the child score less the parent score.

The key finding was that the parents in the clinic group consistently reported more externalizing problems than were reported by the children themselves. This finding applied for both mothers and fathers, and for the male and female children. In addition, the difference between the parent and child externalizing scores was significantly greater in the clinic group than in the community group.

The difference between the total number of problems reported by children and their parents was consistently smaller in the clinic group than in the community group. This finding applied for reports obtained from both mothers and fathers. The smaller difference between parent and child scores in the clinic group arose mainly because the parent scores varied more across the two groups than the child scores. The difference between the parent and child internalizing scores did not vary significantly across the clinic and community groups. This was because the parent-reported and child-reported internalizing scores showed parallel variations across the groups. As a result, the size of the difference between the parent and child internalizing scores was similar in both groups.

There are at least two possible explanations for the difference between the pattern of discrepancies between parent and child scores in the community and clinic groups. First, it is possible that the pattern of discrepancies identified with the clinic-referred children might have been present in all families where parents reported that their children had a large number of problems, regardless of whether or not the children

had been referred to a mental health clinic. Alternatively, it is possible that the pattern found in the reports describing the clinic-referred children reflected specifically the type of childhood problems referred to mental health clinics. In order to investigate this issue further, all the children living in the community who participated in the mail survey were divided into two groups. One group consisted of the children who scored below the recommended total behaviour problem cutoff score on their mother's child behavior checklist (Achenbach & Edelbrock, 1983, p. 63). The other group consisted of children who scored above the recommended cutoff score on their mother's checklist.

The parent and child scores describing the children who scored above or below the cutoff scores on their mother's checklist are shown in Tables 7.7 and 7.8. It can be seen that the pattern of discrepancies between the parent and child externalizing scores in the two groups was similar to that described earlier for the groups consisting of clinic-referred and community children (Tables 7.5 & 7.6). In particular, the parents of children who scored above the cutoff score consistently reported more externalizing problems than the children. As well, the discrepancy between the parent and child externalizing scores was significantly greater in the group containing children who scored above the cutoff score than in the group containing children who scored below the cutoff score.

Three of the male children and two of the female children in the mail survey who scored above the cutoff score on their mother's checklist had attended a mental health clinic in the six months prior to the study (not necessarily one of the clinics participating in the study of clinic-referred children). One of the female children and none of the male

children who scored below the cutoff score on their mother's checklist had attended a mental health clinic during the same period. The scores from the children in the mail survey who had attended a mental health clinic were retained for the purpose of the analyses described above.

The findings from the children who participated in the mail survey suggest that the pattern of discrepancies between parent and child reports described earlier for the clinic-referred children, reflect the pattern of parent and child discrepancies in families in which a parent reports a large number of problems in his or her child. It seems less likely that the results solely reflect a difference between reports describing clinic-referred children and children living in the community. However, the clinic-referred children in this study were all living in two-parent families. In light of this, it is necessary to be cautious in generalizing the findings to children living in single-parent families.

It should be noted that the results presented in Tables 7.5 to 7.8 were based on the standard items comprising the scales on the Child Behavior Checklist and Youth Self-Report. As discussed in Chapter 4, there are some differences in the items which comprise the scales on the two checklists. This makes it difficult to interpret the meaning of differences identified between parent and child scores. In order to address this problem, the modified scales described in Chapter 4 were used to compare more specifically the total number of behaviour problems, and the number of externalizing problems and internalizing problems reported by parents and children in the community and clinic groups. It will be recalled that the modified total problem score was comprised of the sum of the scores from those items which are present on both the checklists. The modified externalizing and internalizing scales were

derived from the externalizing and internalizing scales used for 12-16 year old males (see Chapter 4). Items not present on both the parent and child checklists were omitted from the modified scales. The modifications allowed direct comparisons to be made with regard to the number of problems reported by parents and children, regardless of the age or sex of the children.

The results of the comparisons between the parent and child scores are reported separately for the clinic and community groups (Tables 7.9 to 7.12). It can be seen that in the clinic group there was a consistent pattern for both parents to report more externalizing problems than the children (Tables 7.9 & 7.10). However, the children in the clinic group consistently reported more internalizing problems than the parents (Tables 7.9 & 7.10). This pattern applied for reports describing both the male and female children but was somewhat more pronounced for reports describing the female children. In contrast, in the community group the children consistently reported more externalizing and internalizing problems than their parents (Tables 7.11 & 7.12). This is the same pattern as that identified in Chapter 4 for reports from the total group of children and parents living in the community.

Another difference between the community and clinic groups which is evident in Tables 7.5 and 7.6 is that in the community group, the parents reported that their children had similar numbers of externalizing and internalizing problems. The pattern for reports describing children in the clinic group was different. In the clinic group, both parents consistently reported that the male and female children had more externalizing problems than internalizing problems. In contrast, the children consistently reported that they had more internalizing problems

than externalizing problems. This pattern was particularly pronounced with the female children where it can be seen that the children reported almost twice as many internalizing problems as externalizing problems. In contrast, the parents consistently reported that the female children had more externalizing problems than internalizing problems. A similar pattern was also evident in the comparison of scores from the parents and the children who scored above or below the cutoff score on the CBCL completed by mothers (Tables 7.7 & 7.8).

Two approaches were used to compare more specifically the number of externalizing and internalizing problems reported by the same informant (mother, father, or child) in the clinic group. First, the modified externalizing and internalizing scales described in Chapter 4 were used to compare the number of externalizing and internalizing problems reported by the same informant. Second, T scores were used to compare the number of externalizing and internalizing problems reported by the same informant. As described in Chapter 4, T scores are available for each scale (Achenbach & Edelbrock, 1983) and are provided to facilitate comparisons between scores reported by the same informant on different scales of the child behaviour checklists.

When the number of externalizing and internalizing problems reported by the same informant in the clinic group were compared, again using the modified scales, the results differed for the male and female children (Table 7.13). For the male children, both parents reported significantly more externalizing problems than internalizing problems. In contrast, there was no significant difference between the number of externalizing and internalizing problems reported by the male children. For the female children, there was no significant difference in the number of

externalizing and internalizing problems reported by the parents. However, the female children reported significantly more internalizing problems than externalizing problems.

Finally, the number of externalizing and internalizing problems reported by individual informants in the clinic group were examined using the T scores from each scale (Table 7.14). Although the differences were less evident than when the comparisons had been based on scores from the modified externalizing and internalizing scales (Table 7.13), the overall pattern was similar. A limitation of using T scores to make comparisons between externalizing and internalizing scores reported by individual informants is that the same T score can cover a range of raw scores on each scale. As well, for different informants the same T score can equate to different raw scores. These limitations explain, in part, why the differences found with the raw scores were less evident when T scores were employed for the comparisons.

The correlations between parent and child scores in the community and clinic groups on the Total Behavior Problem Scale, Externalizing Scale, and Internalizing Scale are shown in Tables 7.15 to 7.17. The pattern of correlations was similar in the community and clinic groups and none of the differences between the correlations from equivalent pairs of informants in the community and clinic-referred families were significant at the level of $p<0.05$ (using r to z transformations to test for the significance of differences).

b) DISC/P Interviews: A comparison of scores obtained in the DISC interviews conducted with parents and children is shown in Tables 7.18 and 7.19. The scores compared included the Total Symptom Score, the

Behaviour/Conduct Score and the Affective/Neurotic Score. The Behaviour/Conduct Score consisted of the sum of the scores from the Conduct Disorder and Oppositional Disorder Scales and the Affective/Neurotic Score consisted of the sum of the scores on the Anxiety and Depressive Scales. The Total Symptom Score consisted of the sum of the scores from all the scales.

The pattern of results obtained with the DISC interviews was similar to that obtained with the child behaviour checklists. However, the variation in the difference between the parent and child externalizing scores in the community and clinic groups was somewhat less pronounced. Similarly, although parents in the clinic group consistently reported more externalizing problems than the clinic-referred children, the difference between the parent and child scores was smaller than was evident on the checklists. Finally, the relative difference in the number of externalizing and internalizing problems reported by parents and children which had been evident with scores on the child behaviour checklists, was not present with scores on the DISC/P. In general, the pattern for the correlations between scores obtained from different informants using the DISC/P (Tables 7.20 to 7.22) was similar to that found with the child behaviour checklists.

7.3 DISCUSSION

The key finding in this chapter was that the pattern of discrepancies between parent and child reports was different for reports describing children referred to mental health clinics and reports describing children living in the general community. In particular, the parents of

children referred to mental health clinics consistently reported more externalizing problems than were reported by their children. This pattern was the reverse of the general pattern found with reports describing the behaviour of children in the community. In reports describing these latter children, the children consistently reported more externalizing and internalizing problems than were reported by their parents.

The results in this chapter provide an explanation for some of the inconsistencies between the results which have been reported in previous studies. As discussed in Chapter 4, in earlier studies children reported that they had more internalizing problems than were reported by their parents. However, parents reported that their children had more externalizing problems than were reported by the children themselves (Herjanic and Reich, 1982; Edelbrock et al., 1986). In contrast, in the study by Offord et al. (1986), the children reported more externalizing problems than their parents. Finally, in this study, children living in the community reported more externalizing and internalizing problems than were reported by their parents (see Chapter 4).

The results in this chapter suggest that part of the explanation for the inconsistencies in earlier studies may be that the studies focused on different populations of children. For example, the studies by Herjanic and Reich (1982), and Edelbrock et al. (1986), included only clinic-referred children while the study by Offord et al. (1986), included only children living in the community. This study included samples of children drawn from both the general community and from children referred to mental health clinics. Consistent with the results of the study by Offord et al. (1986), in this study children living in the community

reported more externalizing and internalizing problems than their parents (see Chapter 4). In contrast, the parents of the clinic-referred male and female children consistently reported more externalizing problems than their children while the children consistently reported more internalizing problems than their parents. This latter pattern is the same as that described in the studies of clinic-referred children by Herjanic and Reich (1982), and Edelbrock et al. (1986).

It should be noted that there were significant differences between the number of externalizing and internalizing problems reported by the same informants across the community and clinic groups, and across the groups defined on the basis of the total problem score reported by mothers. This suggests that there was general agreement amongst parents and children that the clinic-referred children, or the children scoring above the recommended cutoff score on their mother's checklist, did have more problems than other children in the community. However, across the groups the pattern of discrepancies between the number of externalizing and internalizing problems reported by the parents and children was different. For the majority of children in the community there appeared to be a general pattern of children reporting somewhat more externalizing and internalizing problems than their parents. In contrast, for clinic-referred children, the parents reported more externalizing problems than were reported by the children.

It is likely that there were a number of factors which contributed to the different pattern of discrepancies between the parent and child reports in the community and clinic groups. For example, it is possible that the parents who reported more externalizing problems in their child were less tolerant of this type of behaviour in their child. A similar suggestion

has been made by Herjanic and Reich (1982) who pointed out that children's symptoms which are reported more frequently by mothers appear to be characterized by "behavior that might be troublesome to the mother" (p. 321). As well, in this study it appeared that the children with externalizing problems who were attending mental health clinics had been punished in the past for episodes of misbehaviour both at home and at school. As a result, it is plausible that these children reported fewer externalizing problems than their parents, not solely because they were unaware of their behaviour problems, but because in the past they had been punished when they had admitted such behaviour.

It is usually a parent or teacher rather than a child, who decides that a child should be referred to a clinic. As such, it seems plausible that children who are referred to clinics are more likely to be those who have problems that are apparent to parents and teachers, particularly children with problems which are irritating or annoying to these people. It seems less likely that a referral will be initiated when only the child is aware of his or her emotional or behavioural problems. This is of particular concern in view of the increasing incidence of suicide amongst young people and the concern that suicide amongst children may be preceded by emotional problems which are not recognized by parents and teachers. It is also of concern because referral to a mental health clinic has been used as a criterion against which the validity of questionnaires assessing childhood emotional and behavioural problems have been tested. For example, this criterion was used to establish the cutoff scores employed with the checklists used in this study (Achenbach & Edelbrock, 1983, 1986, 1987). The results in this chapter suggest that although children referred to mental health clinics may have more problems than children in the general community, the extent and nature of

these problems appear to be perceived differently by parents and children. Thus, caution is needed when using referral to a mental health clinic as a criterion for childhood psychological problems.

The results reported in this chapter extend the earlier work by Mokros et al. (1987). In their study, Mokros et al. (1987) focused on symptoms associated with childhood depression which were grouped into "behavioural" and "ideational" categories. In both areas, children referred to a mental health clinic reported fewer problems than their parents while children living in the general community reported more problems than their parents. However, as discussed in Chapter 1, the study by Mokros et al. (1987) included only a small sample of female children, the "parents" of the clinic-referred children included both mothers, fathers, and grandparents, and the symptoms studied were limited to those of childhood depression. In the present study, reports about externalizing and internalizing problems covered a wide range of areas and separate reports were obtained from both parents, and from male and female children. In addition, the results were analyzed separately for the mothers and fathers, and for the male and female children. This approach made it possible to see that a somewhat different pattern of results applied with different pairs of informants when describing male and female clinic-referred children. In particular, differences between the number of internalizing problems reported by parents and the female children were substantially larger than the differences between parents and the male children. It is likely that the different ages of children in the two studies, the different symptoms studied, and the small number of female children in the clinic sample, explain the somewhat different results found in the study by Mokros et al (1987).

The number of community and clinic-referred families who participated in the DISC/P interviews was smaller than the number who completed the child behavior checklists. As well, for the purpose of the study only items comprising the conduct disorder, oppositional disorder, depression, and anxiety scales were included in the interviews. The items from the first two scales made up the "Externalizing" scale in the DISC/P interviews. Items from the latter two scales made up the "Internalizing" scale in the interviews. The more narrow focus of the DISC/P interviews may explain, in part, why there were some differences between the results obtained with the DISC and DISCP, and with the child behavior checklists, particularly in the area of externalizing problems. The externalizing items which comprised the DISC/P scales tended to represent two extremes. The items included from the conduct disorder scale tended to be severe and were not often endorsed by parents and children in either the community or clinic-referred families. The items included from the oppositional disorder scale tended to be symptoms which applied more for younger children and were less common amongst the young adolescents. In light of this, it is possible that there was less opportunity for parents and children in the community and clinic groups to describe externalizing problems during the interviews. As a result, the varying pattern of externalizing scores evident when reports were obtained using the child behaviour checklists, were less evident when reports were obtained using DISC/P interviews.

Overall, it appears that children in the community generally report more internalizing and externalizing problems than their parents. However, in circumstances where a child has been referred to a mental health clinic, parents report more externalizing problems than their child. In these circumstances it appears that reports from parents will highlight the

presence of externalizing problems in their child, however, reports from the child will draw attention to the child's internalizing problems. The findings highlight the limitations of studying only clinic-referred populations of children and also emphasize the limitations of using referral to a mental health clinic as a criterion for childhood emotional and behavioural problems. Children referred to mental health clinics form only a small percentage of the total population of children in the community (Sawyer et al., in press) and it appears that reports of their behaviour problems obtained from different informants may not reflect patterns which may apply for the much larger population of children living in the general community.

CHAPTER 8. A COMPARISON OF REPORTS DESCRIBING CHILDHOOD EMOTIONAL AND BEHAVIOURAL PROBLEMS REPORTED DURING INTERVIEWS WITH MOTHERS, FATHERS, AND CHILDREN

8.0 INTRODUCTION

Mail surveys are a well established method of collecting data in other fields (Baghurst et al., 1987a; Baghurst et al., 1987b; Dillman, 1978), however, they have not been widely employed for data collection in child psychiatry. As well, although behaviour checklists are widely used for data collection in studies of childhood emotional and behavioural problems, clinicians generally obtain information about children by means of interviews with parents and children. A possible concern about using a mail survey and behaviour checklists for data collection in this study is that the results may differ from those which would be found if information had been collected by means of interviews with parents and children. If so, this would reduce the relevance of the findings of the study for clinicians working in the area of child psychiatry.

There are two reasons why the results in this study may have differed from those which would have been found if all the information about the children had been collected by means of interviews. First, the results obtained using behaviour checklists may not reflect the results which would be obtained if information about children's behaviour was collected by means of interviews. Second, the use of a mail survey in the study left open the possibility that parents and children collaborated when

answering their questionnaires rather than completing them independently. If this had occurred, it may have biased the results obtained in the mail survey.

This chapter focuses on these two methodological issues. The first part of the chapter examined whether the same pattern of differences between parent and child reports existed when information was obtained by means of separate interviews with parents and children, rather than by using behaviour checklists in a mail survey. The second part of the chapter investigated whether the pattern of results obtained on behaviour checklists completed during a home visit, which were known to have been completed independently by parents and children, was the same as that obtained on the checklists completed during the mail survey.

8.1 METHOD

8.11 Subjects

The subjects in this chapter were the same sub-sample of community parents and children who were described in Chapter 7.

8.12 Procedure

The DISC interviews and the checklists were all administered during a single home visit. As mentioned in Chapter 7, although it was not possible to randomize completely the order in which the interviews and checklists were administered, this was done as far as was practicable.

8.13 Measures

Full details of the measures employed in this chapter are provided in Chapters 2 and 7. As described in these earlier chapters, for the purpose of this study only items from the conduct disorder, oppositional disorder, anxiety, and depressive areas of the DISC/P were included in the interviews conducted with parents and children. This was done to keep the total length of the DISC/P interviews to manageable proportions. The content of the items in these areas is summarized in Table 8.1.

8.14 Data Analysis

Analyses reported in this chapter focused on two methodological issues. The first set of analyses focused on whether the same pattern of differences between parent and child reports existed when information was collected by means of separate interviews with parents and children rather than by means of the child behaviour checklists. The second set of analyses tested whether the difference between parent and child scores on the child behaviour checklists was less in the mail survey than when the same checklists were completed during a home visit. The latter analyses provided a check on whether parents and children independently completed the checklists during the mail survey.

Once again, a non-parametric statistic (the Wilcoxon Signed-Ranks Test) was used to test the significance of differences between scores obtained from the parents and children. The significance of the differences in the discrepancies between parent and child scores across the groups, which showed an approximately normal distribution, was tested using t-tests. Finally, Pearson product-moment correlations were used to test

the consistency of scores obtained from different informants during the home visits and in the mail survey.

8.2 COMPARISON OF SCORES OBTAINED FROM DISC AND DISCP INTERVIEWS

8.21 Comparison of Mother and Father DISCP Scores

The mean scores obtained from the mothers and fathers of the male and female children are reported in tables 8.2 and 8.3.

Although the number of symptoms reported in some areas was small, the overall pattern of results was the same as that found with the child behaviour checklists in the mail survey. In particular, on all the scales the fathers of male and female children consistently reported fewer symptoms than the mothers. The difference between parent total symptom scores and behaviour/conduct scores was significant at the level of $p<0.05$ for both the male and female children. Differences between the parents for affective/neurotic scores were significant for the female children, but not for the male children, at the level of $p<0.05$.

8.22 Comparison of Parent and Child DISC/P Scores

The mean scores obtained from the parents and children are shown in tables 8.4 to 8.7.

Once again, the pattern of scores on the DISC/P was similar to that which had been found with the child behaviour checklists. In particular, the scores reported by the children on the total symptom scale,

behaviour/conduct scale, and the affective/neurotic scale were consistently higher than the scores reported by the parents. The latter was particularly evident for reports describing the female children. Except for the mothers and male children, all the differences between parent and child total symptom scores were significant at the level of $p<0.05$. The differences between the affective/neurotic scores reported by the male and female children and the parents were all significant at the level of $p<0.05$. However, only the difference between fathers and the male children on the behaviour/conduct score achieved this level of significance.

The pattern of affective/neurotic scores reported by children in the DISC interviews appeared to differ from the pattern reported on the child behaviour checklists. Specifically, in the DISC interviews, female children reported significantly more affective/neurotic symptoms than the male children ($t=-3.7$, $df=79$, $p<0.0004$). Although female children reported more internalizing problems on the checklists than the male children, the difference was relatively small.

When scores from the checklists completed during the mail survey by the male and female children who participated in the DISC interviews were re-analysed, it was found that the female children had also reported significantly more internalizing problems than the male children in the mail survey ($t=-3.0$, $df=79$, $p<0.004$). In addition, this particular group of male children had reported somewhat fewer internalizing problems ($t=1.8$, $df=251$, $p<0.07$) than the other male children who participated in the mail survey. One possible explanation for the latter result was that the male children in two-parent families, from whom the sub-sample of children were selected for DISC interview, reported fewer internalizing

symptoms than male children in single-parent families. If true, the exclusion of the latter children might have resulted in a lower mean score for the remaining children. However, when this possibility was tested, it was found that there were only small differences between the internalizing scores reported by the male children in the single-parent and two-parent families. Instead, it appears that by chance the group of male children randomly selected for DISC interviews had fewer internalizing or Affective/Neurotic symptoms than the total group of male subjects who participated in the mail survey.

8.23 Correlations between Parent and Child DISC/P Scores

The size of the correlations between the total symptom scores, behaviour/conduct scores, and affective/neurotic scores reported by mothers, fathers, and children are shown in Tables 8.8 to 8.10. To facilitate comparisons with earlier studies, Pearson Correlations are reported, however, since these may be sensitive to pairs of observations lying distant from the majority, rank correlations were also estimated. The results of the rank correlations varied little from those obtained using the Pearson Correlations.

The pattern of the estimated correlations was generally similar to that obtained with scores from the child behaviour checklists. However, the estimated correlations between the mother and child scores on the Affective/Neurotic scale of the DISC were consistently higher than the correlations between the mother and father scores. This pattern was also present between the total symptom scores describing the male children. In contrast, the correlations between behaviour/conduct scores followed

the same pattern as that identified earlier with scores from the child behaviour checklists.

8.3 CHILD BEHAVIOUR CHECKLIST SCORES OBTAINED DURING HOME VISITS

8.31 Comparison of Child Behaviour Checklist scores obtained in the Mail Survey and scores obtained during the Home Visits

The mean difference between the parent and child scores reported during the mail survey, and the mean difference between parent and child scores reported during the home visits are shown in Table 8.11. It can be seen that the discrepancy between the scores reported by parents and children varied little across the groups and none of the variations were significant at the level of $p<0.05$.

The scores obtained from parents and children during the mail survey and subsequently during the home visit are shown in Table 8.12. In contrast to the lack of variation in the discrepancies between the informant scores reported in the mail survey and during the home visits, it can be seen that the scores reported by the same informants during the home visit were consistently lower than those reported in the earlier mail survey. This pattern was particularly evident for reports describing the male children.

8.32 Correlations between Child Behaviour Checklist scores reported in the Mail Survey and scores reported during the Home Visits

The size of the correlations between parent and child scores obtained in the mail survey and between scores reported on the checklists completed during the home visits are shown in Tables 8.13 to 8.15.

The size of the correlations between the scores reported by different informants in the mail survey and between the scores reported during home visits were similar and none of the differences between the correlations found in the mail survey and those between scores reported during the home visits were significant at the level of $p<0.05$ (using r to z transformations to test for the significance of differences). However, it can be seen that the correlation between the mother and father total problem scores, and between the parent and child internalizing scores reported during the home visits, were somewhat smaller than the equivalent correlations obtained in the mail survey. In contrast, the correlations between the parent and child total problem scores and the parent and child externalizing scores reported during the home visit were somewhat larger than the equivalent correlations obtained in the mail survey.

8.4 DISCUSSION

In general, the similarity of the results obtained from the behaviour checklists completed during the mail survey, and those completed during the home visits suggests that parents and children independently completed the checklists during the mail survey. The result was also consistent with the impression gained by the research assistant who undertook the home visits. During the home visits, the research assistant took the opportunity to discuss the mail survey with participating families. These conversations strongly suggested that family members had independently completed their questionnaires as requested. In a small minority of families, parents had compared their answers after completing their questionnaires. However, the right of the children to complete their questionnaires in private had been respected.

There are at least three possible explanations for the decrease in behaviour checklist scores reported by parents and children at the time of the home visit. One possibility is that parents and children report fewer problems during a home visit when a research assistant is present, than when the same information is collected in a mail survey. An alternative explanation is that the decreased scores reflect a decrease in the number of childhood problems amongst children in the study. Finally it is possible that the lower scores are a "practice effect" arising as a result of the informants having previously completed the same checklist. A decrease in the number of problems reported by parents and children at the time of their second interview was reported by Edelbrock et al. (1985). The phenomenon appears to be common when

questionnaires or interviews are administered on a second occasion and it seems possible that in this study, the decline in scores at the second assessment was primarily a reflection of this phenomenon.

It should be noted that it was not feasible, with the methodology used in this study to investigate which was the most likely explanation for the decline in checklist scores reported during the home visits. To do this it would be necessary, at a minimum, to randomize the order in which the two data collection procedures were administered. In this way it would be possible to determine whether the differences reflected the type of methodology used for data collection, or whether they reflected an effect of repeated administration of the checklists. However, the finding highlights the importance of including control groups when testing the effectiveness of interventions aiming to decrease the prevalence of childhood emotional and behavioural problems. Failure to include control groups could lead to an interpretation that a decrease in the number of problems reported by parents and children was due to the beneficial effects of an intervention. Instead, the decrease may reflect the effects of repeated administration of the same interviews or questionnaires.

The pattern of scores reported in the DISC and DISCP interviews by the parents and children was similar to those reported on the checklists. This is consistent with the results of the study by Kazdin, Esveldt-Dawson, Unis, and Rancurello (1983) which reported that the correspondence between parent and child reports did not vary as a function of the assessment format employed to obtain the reports from the children and parents (self-report or interview). Ivems and Rehm (1988) have commented that when the same rater is used to interview both

children and parents, better levels of agreement may be found between the two informants. This may provide a partial explanation for why some of the differences between parent and child scores appeared smaller when information was collected by means of the DISC/P interviews. As well (see discussion in Chapter 7), the range of externalizing items in the DISC/P interviews used in this study was more restricted than that which applied with the child behaviour checklists and this may have contributed to the smaller differences between the DISC/P scores in this area.

In summary, the results obtained using the child behaviour checklists and the DISC/P interviews were similar. As well, the patterns of informant scores on the checklists completed during the home interviews were similar to those found in the mail survey. The general consistency of results obtained using the different approaches for data collection suggests that the results in this study are not an artifact of a particular method of data collection, but reflect genuine differences between the pattern of reports from parents and children describing the problems of children living in the community or attending mental health clinics.

CHAPTER 9. SUMMARY

This chapter summarizes the key findings from the study, the implications of the findings for the assessment of childhood emotional and behavioural disorders, and possible directions for future research.

9.0 THE CONTEXT OF THIS STUDY

The diagnostic assessment of childhood emotional and behavioural disorders requires the evaluation of children's behaviour in a range of different settings. In order to achieve this, clinicians and researchers have to rely on reports from parents, teachers, and children themselves. During the last several years, there has been increasing evidence that there is poor agreement between the reports obtained from these informants (Achenbach et al., 1987). This is important because differences in diagnosis and formulation will inevitably lead to different treatment plans for children.

Currently, there are no guidelines available to advise clinicians or researchers as to how they should weight discrepant reports obtained from different informants. As a result, assessments of childhood disorders will vary, depending on the choice of informants, the extent of agreement between informants, and whether reports from one informant are weighted more heavily when disagreement exists. A better understanding of the extent of the differences between informant reports and the reasons for these differences will help improve the accuracy of assessment and

formulation, and hence the quality of treatment for children. A better understanding of these differences is also essential before methods for combining reports from different informants can be developed.

9.1 THE AIMS AND FINDINGS OF THIS STUDY

The principal aim of this thesis was to study the level of agreement between parents, children, and teachers when describing childhood emotional and behavioural problems. In addition, the thesis investigated plausible reasons which may explain why discrepancies exist between the reports provided by these informants. Finally, the pattern of discrepancies between parent and child reports describing different populations of children were compared.

The unique features of the study included the comparison of reports from four informants (mothers, fathers, teachers, and children) describing the same group of children, the use of similar measures to obtain reports from the different informants, the inclusion of children in an age range where the reliability of childhood reports is acceptable (Edelbrock et al., 1985), and the comparison of discrepancies between informant reports in two different populations of children.

For the children living in the general community, when the number of childhood problems reported by different informants was compared (Chapter 4), a clear pattern was evident. Children consistently reported the most problems, teachers reported the fewest problems, and the number of problems reported by parents was intermediate between those reported by children and teachers. For children in the general community, this

pattern applied for reports describing both externalizing and internalizing problems.

Several possible reasons which may explain why the children in the community reported more behaviour problems than the parents or teachers were discussed. For example, it is possible that the children have a lower threshold for reporting problems. Alternatively, children's knowledge of their behaviour in a wider range of settings than other informants may result in them knowing about more problems than are known by other informants. It is plausible that teachers report fewer problems than the other informants because they have more limited opportunities to observe children's behaviour. As well, teachers may have a higher threshold for reporting problems. Finally, teachers may be concerned that acknowledging the presence of large numbers of problems amongst students in their class may reflect badly on their ability as a teacher.

Although there was a large difference in the number of problems reported by different informants, the correlations between externalizing and internalizing scores reported by different informants were somewhat higher than has been reported in earlier studies (Achenbach et al., 1987). The similarity of the questionnaires used by all the informants in this study may, in part, explain why the correlations were higher than those reported in previous investigations.

A comparison of the number of children identified as "cases" by different informants, on the basis of their exceeding the recommended cut-off score on the Child Behavior Checklist, Teacher Report Form, or Youth Self Report Form (Achenbach & Edelbrock, 1983, 1986, 1987) revealed a number of important findings (Chapter 5). First, the number of cases varied,

depending on which informants were employed to describe the children's behaviour. For example, when reports from all four informants were examined, a total of twenty-six 10-11 year old female children were identified as cases. If reports from only the mothers and teachers had been examined, only twenty of these children would have been identified as cases. This was because six children were identified as cases only on the father and/or child reports. Further analysis revealed that approximately 50% of the children identified as cases were identified as a case by only one informant. The results of the study suggested that the degree of overlap between cases identified by the different informants was no better than might be expected by chance.

When the level of agreement between informants about the presence of individual childhood emotional and behavioural problems was assessed (Chapter 5), the results were similar to those described in earlier studies (Herjanic & Reich, 1982). The problems for which there was the highest agreement were generally those which would be apparent to an external observer. These included problems such as "Poor schoolwork", "Bites fingernails", and "Gets in many fights". In contrast, the problems showing the lowest level of agreement were generally problems that would be less obvious to external observers such as "Strange ideas", "Compulsions" and "Feels too guilty". When the prevalence of individual problems as reported by different informants was compared, a very consistent pattern was evident. Children reported the highest prevalence, teachers reported the lowest prevalence, and mothers and fathers reported a prevalence which lay between that of the children and teachers.

The apparent influence of three broad groups of factors on the level of agreement between parents and children was investigated (Chapter 6). The factors included demographic characteristics of the parents, parent psychopathology, and the psychological adjustment of the family. Overall, the demographic characteristics of the parents did not have a large influence on the level of agreement between parents and children. However, in some instances differences in the demographic characteristics of informants were associated with differences in reports from one informant when describing children's behaviour, but not another. For example, mothers with less education reported significantly more problems in their children than were reported by mothers who had achieved higher levels of education. In contrast, there were no significant variations in the child scores across the groups defined on the basis of maternal education. This finding raises the possibility that informants may differ not only in their perception of children's behaviour, but also in their perception of factors which may impinge on children's behaviour. If this occurred consistently, it might explain why reports from different informants revealed different relationships between children's behaviour and other factors in the child's environment.

The apparent effect of parent psychopathology on parent reports varied for mothers and fathers, and for reports describing children of different sex (Chapter 6). For example, the positive association between self-reported psychopathology and mothers' reports describing the problems of the male children was stronger than the equivalent association between self-reported psychopathology and mothers' reports describing the problems of the female children. When reports from different parents in two-parent families were compared, it was found that "disturbed" parents (defined on the basis that the parent self-report score exceeded the

recommended cutoff on the General Health Questionnaire) consistently reported more problems in children than their "healthy" spouse. The effect was evident regardless of the sex of the parent but was more pronounced for reports describing the male children. There was also a significant positive relationship between parent perception of their family's adjustment and their perception of behaviour problems in their child. Once again, the strength of the relationship varied for parents and children of different sex. There are a number of possible reasons for these findings. For example, it is plausible that children exhibit more behaviour problems with parents who have psychological problems. Alternatively, parents who perceive themselves to have more psychological problems may also perceive more problems in their child. This might occur because distressed parents project some of their own distress onto their child.

The results of analyses which focused on the level of agreement between parent and child reports, taking into account parent psychopathology and family adjustment, differed for mothers and fathers when describing children of different sex (Chapter 6). Not surprisingly, the amount of variance in scores from the older male children accounted for by scores from their parents was less than that which applied with the younger male children. Older children often spend a considerable amount of time away from their family home and it is possible that, as a result, parents of older children may be less familiar with the behaviour of their children. The results for the female children showed a different pattern. For example, the amount of variance in the scores from the older female children accounted for by the mother's score was similar to that which applied with the younger children. In contrast, father scores accounted for a particularly low proportion of the variance in the scores from the

older females. It is possible that this reflects a better level of communication between mothers and their older daughters than occurs between the daughters and their fathers.

The pattern of discrepancies between parent and child reports was different for the children referred to mental health clinics and the children living in the general community (Chapter 7). In particular, the parents of the clinic-referred children consistently reported more externalizing problems than the children, a result which was not evident for the majority of children living in the community. As well, the difference between parent and child externalizing scores was significantly greater for the clinic-referred children than for children in the community. The pattern found with the clinic-referred children was also found with children in the community whose mothers reported that their child had a large number of problems. There are several possible reasons for the different pattern of reports found with the community and clinic-referred children. For example, the parents of children referred to mental health clinics may be less tolerant of their children's behaviour, particularly behaviour which irritates or annoys the parents. It is also possible that in the past, the clinic-referred children have been punished when they admitted misbehaving. As a result, they may tend to deny the presence of externalizing behaviour problems.

Finally, the pattern of results found when information was collected by means of separate interviews with parents and children, was similar to that found using behaviour checklists in the mail survey (Chapter 8). As well, when parents and children completed the behaviour checklists during a home visit by a research assistant, the results were essentially the same as those found with the behaviour checklists completed during the

mail survey. These findings suggest that the results obtained in the mail survey were not an artifact of a particular method of data collection, but reflected genuine differences between the pattern of reports describing children's behaviour obtained from different informants.

9.2 THE IMPLICATIONS OF THE RESULTS OF THIS STUDY

The results of this study have significant implications for work in a number of areas relevant to the assessment and treatment of children with emotional and behavioural problems.

First, the results have implications for the clinical assessment of children with emotional and behavioural problems. For example, the majority of children in the community reported more externalizing and internalizing problems than their parents or teachers. However, the pattern for clinic-referred children, or children whose parents reported that their child had a large number of problems, was different. Parents of these children reported that the children had more externalizing problems than were reported by the children. It appeared that for clinic-referred children, reports from parents were highlighting the presence of externalizing problems, while reports from the children drew more attention to the presence of internalizing problems. The limited agreement between reports obtained from parents and children was also highlighted by the relatively low correlations between scores from the different informants. At best, scores reported by parents accounted for only 41% of the variance in scores reported by children, even after parent psychopathology and family adjustment were taken into account. In some instances, such as reports from fathers describing older female

children, the variance in the child score accounted for by the parent score was much lower.

The difference between reports from different informants poses a considerable difficulty for clinicians who combine reports from different informants in order establish a child's diagnosis. The relatively low correlations between scores also suggests that clinicians need to be cautious when substituting parent or teacher reports for reports from children. It seems likely that for the majority of children in the community, if greater weight is given to the reports from children, a larger number of childhood problems will be identified. If greater weight is given to reports from parents and teachers, fewer problems will be recognized. The situation with clinic-referred children, or children whose parents report that their child has a large number of problems, may be somewhat different. Overall, like their parents, these children report more problems than other children. However, the reports from the children appear to present a somewhat different picture of children's problems than the reports from the parents. In particular, reports from the children draw more attention to the presence of internalizing problems. In contrast, reports from the parents emphasize the presence of externalizing problems. These findings have particular relevance for the clinical assessment of children referred to mental health clinics. For these children, the results suggest that the use of reports from different informants may influence not only whether or not problems are identified, but also the conclusions which are reached about the nature of the problems. For example, if greater weight is given to reports from parents, a clinician may be more likely to diagnose the presence of a childhood conduct disorder or externalizing problem. In contrast, if greater weight is given to the child's report, it appears more likely

that a clinician will diagnose an internalizing problem. If equal weighting is given to all reports, a clinician may conclude that the child warrants more than one diagnosis, for example, conduct disorder and depression. These suggestions are consistent with the results from the studies by Reich et al. (1982) and Kashini et al. (1985). In the study by Reich et al. (1982), depression was found to be the most common disorder when diagnoses were based on children's reports of their own problems, and attention deficit disorder and oppositional disorder were the most common disorders when diagnoses were based on parental report. Similarly, Kashini et al. (1985) noted that a majority of children diagnosed with depression on the basis of the children's reports "were considered by parents to have attention deficit or oppositional disorder" (p. 441).

The two major systems used to classify children's emotional and behavioural disorders (American Psychiatric Association, 1980; World Health Organization, 1978) provide little guidance with regard to how children's disorders should be correctly diagnosed when there are differences between the reports provided by various informants. Yet the possibility that different diagnoses may be assigned, depending upon the weighting given to different reports, poses significant problems for the valid and reliable diagnostic assessment of children. This is important because differences in diagnosis and formulation inevitably lead to different treatment plans for children with emotional and behavioural problems. At present, it appears that clinicians need to be particularly sensitive to the possibility that information about clinic-referred children obtained from parents may emphasize the presence of externalizing problems while reports from children will draw more attention to the presence of internalizing problems.

Second, the results of the study have implications for the use of "referral to a mental health clinic" as a "gold standard" against which the significance of reports of problems from children and other informants can be measured. Attempts to identify the clinical significance of reports from different informants describing childhood emotional and behavioural problems have been hindered by the lack of an accepted standard against which the clinical significance of the reports can be tested. One approach which has been employed to address this problem is to use "referral to a mental health service" as a criterion of children's need for help (Achenbach & Edelbrock, 1983, 1986, 1987). However, the results of this study suggest that children referred to mental health clinics may be a particular sub-group of the total group of children with problems in the community. In particular, clinic-referred children appear to be characterized by parent reports which emphasize the presence of externalizing problems in the children. A possible explanation for this result is that children referred to clinics are more likely to be those who have problems which are apparent to parents, particularly problems which are irritating or annoying for parents. In light of this finding, it may be more accurate to use "referral to a mental health service" as a criterion of children's need for help, as recognized by parents. What is of particular concern is the possibility that problems which primarily cause distress for children, but which may be less obvious to parents or teachers, may not initiate referral to mental health services. As a result, referral to a mental health clinic may not be an appropriate criterion to use when determining the significance of reports from children. Instead, alternative criteria may be needed to establish the clinical significance of reports from children. These criteria could include the prognosis of problems

identified specifically by children and the impact of such problems on the overall development of children.

Finally, the results of this study have important implications for epidemiological studies which use a two-stage sampling procedure in order to identify children with problems. It appears that when cutoff scores on checklists are used to identify the initial sample of children, different children ("cases") will be identified by different informants. In fact, the extent of overlap between "cases" identified by different informants may be no larger than would be expected by chance. This is an important finding because children identified by different informants may have different characteristics. If so, studies which rely solely on reports from mothers, for example, will not provide a complete picture of the problems suffered by all children.

Obtaining reports from several informants in large epidemiological studies is clearly time consuming and expensive. It is possible that such studies will need to set specific goals and select informants whose reports are relevant to these goals. For example, it appears that studies investigating the relationship between parental psychiatric disorders and childhood emotional and behavioural disorders may find a somewhat higher prevalence of childhood problems if reports are obtained only from the parent with psychological problems. In light of this, obtaining reports from more than one parent may be an important element of such studies.

9.3 DIRECTIONS FOR FUTURE RESEARCH

The findings from this study suggest three broad areas for future investigation.

First, the results of the study highlight the extent of differences between reports obtained from parents, teachers, and children in the general community. The results also draw attention to the apparent influence of factors such as parent psychopathology and family adjustment on reports from different informants. To date, however, studies in this area have employed cross-sectional designs. Although these designs are useful in order to demonstrate associations between important variables, prospective studies are needed to demonstrate the cause and effect relationships between these variables. For example, prospective studies could provide information about whether a deterioration in a parent's psychological adjustment is associated with a subsequent increase in the difference between a parent's reports of their child's behaviour and reports from other informants. Alternatively, changes in family functioning or a child's age may be associated with subsequent changes in the level of agreement between parents and children.

Second, there is a need to investigate the clinical significance of reports about childhood emotional and behavioural problems obtained from different informants. At present, a major difficulty in determining the clinical significance of problems reported by only one informant, particularly when the informant is a child, is that there is little information available about the prognosis of such problems. For example, it is possible that problems identified solely by a child are more

transient in nature than problems identified by both the child and their parents. Alternatively, a problem reported by a child but not by their parents may have added significance in some situations. For example, suicidal thoughts reported by a child may be a much more accurate predictor of childhood suicide than reports from parents. Prospective studies are needed to investigate the relationship between reports from specific informants and the later development of childhood or adult disorders.

Third, the extent and clinical significance of "informant specific" relationships between various factors and childhood emotional and behavioural problems needs further study. This is particularly relevant for reports obtained from children. In the past, it had been assumed that relationships between children's behaviour problems and factors "external" to children's behaviour would be the same, regardless of the informant used to describe the children's problems. Results from this study suggest that this assumption may not be correct. It is possible that reports from different informants may reveal different relationships which are relevant to the aetiology, prognosis, or treatment of childhood disorders. For example, it is possible that some children become unhappy or depressed when they perceive themselves to have a poor relationship with their parents, despite the fact that the parents fail to recognize this sequence of events. If this happened consistently, a different relationship between problems with parents and childhood depression might be identified if children rather than parents were used as informants about childhood problems. At present, it is difficult to determine the significance of relationships between emotional and behavioural problems and external factors when they are identified by children. This is because most current information about relationships between children's

behaviour and external factors has been based on reports from parents. Further investigation of relationships identified using reports from different informants may suggest new interventions which could be used to prevent the onset of childhood emotional and behavioural problems, or provide more adequate help once problems have already developed.

Finally, the results of the study draw attention to the complexities inherent in combining reports from different informants when assessing childhood emotional and behavioural problems. Despite these apparent complexities, this is the approach routinely taken by clinicians when completing the diagnostic assessment of a child. There is a great need to investigate further the extent to which it is possible to combine reports from different informants to produce a valid and reliable assessment of childhood emotional and behavioural problems.