

Bangor University

PROFESSIONAL DOCTORATES

The teddy bear's picnic : a study comparing 5 year old children's responses to a new narrative measure, with parent and teacher rated psychological problems and global concern.

Hammond-Rowley, Sara

Award date:
1999

Awarding institution:
Bangor University

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

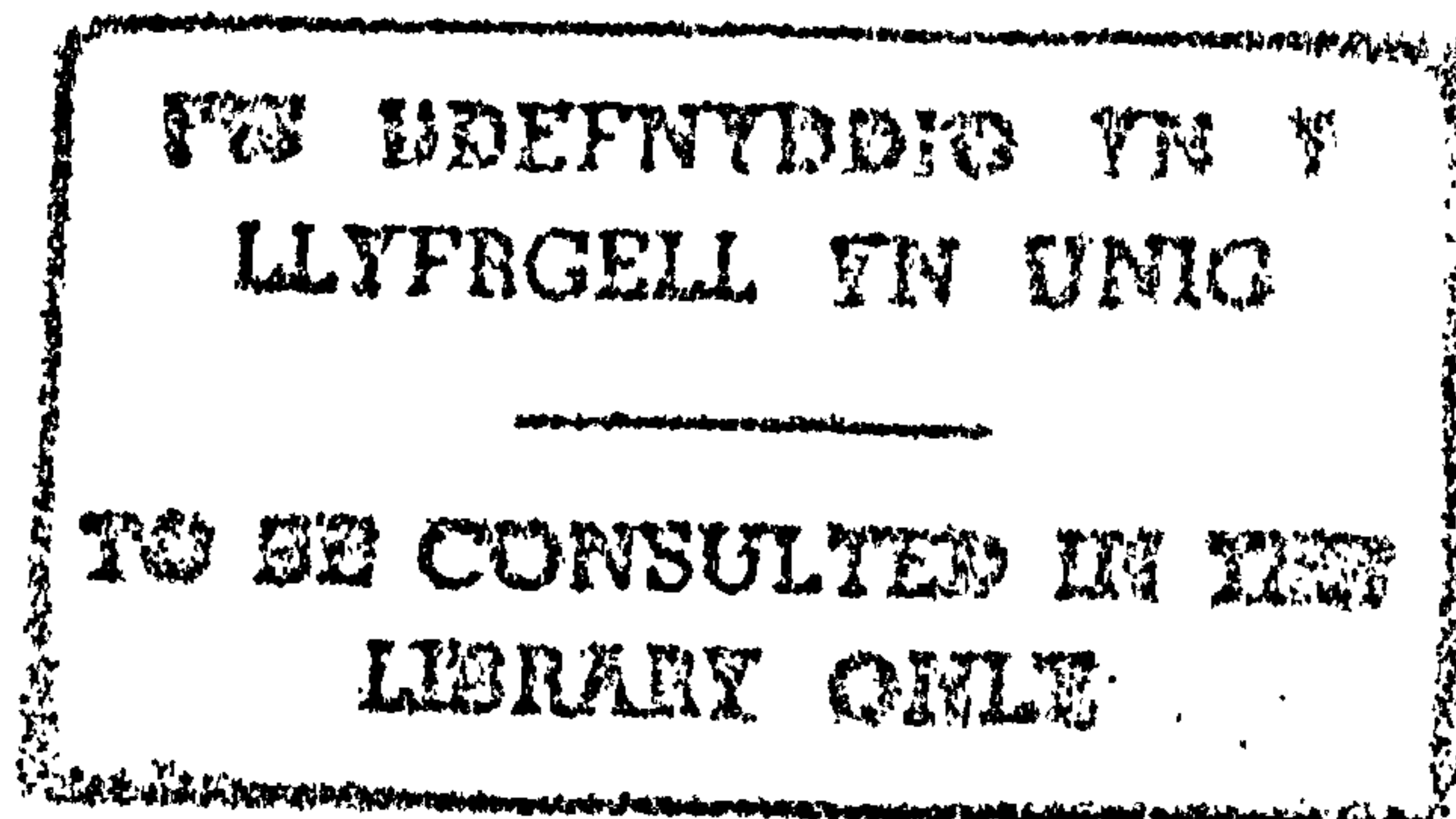
Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The Teddy Bear's Picnic:

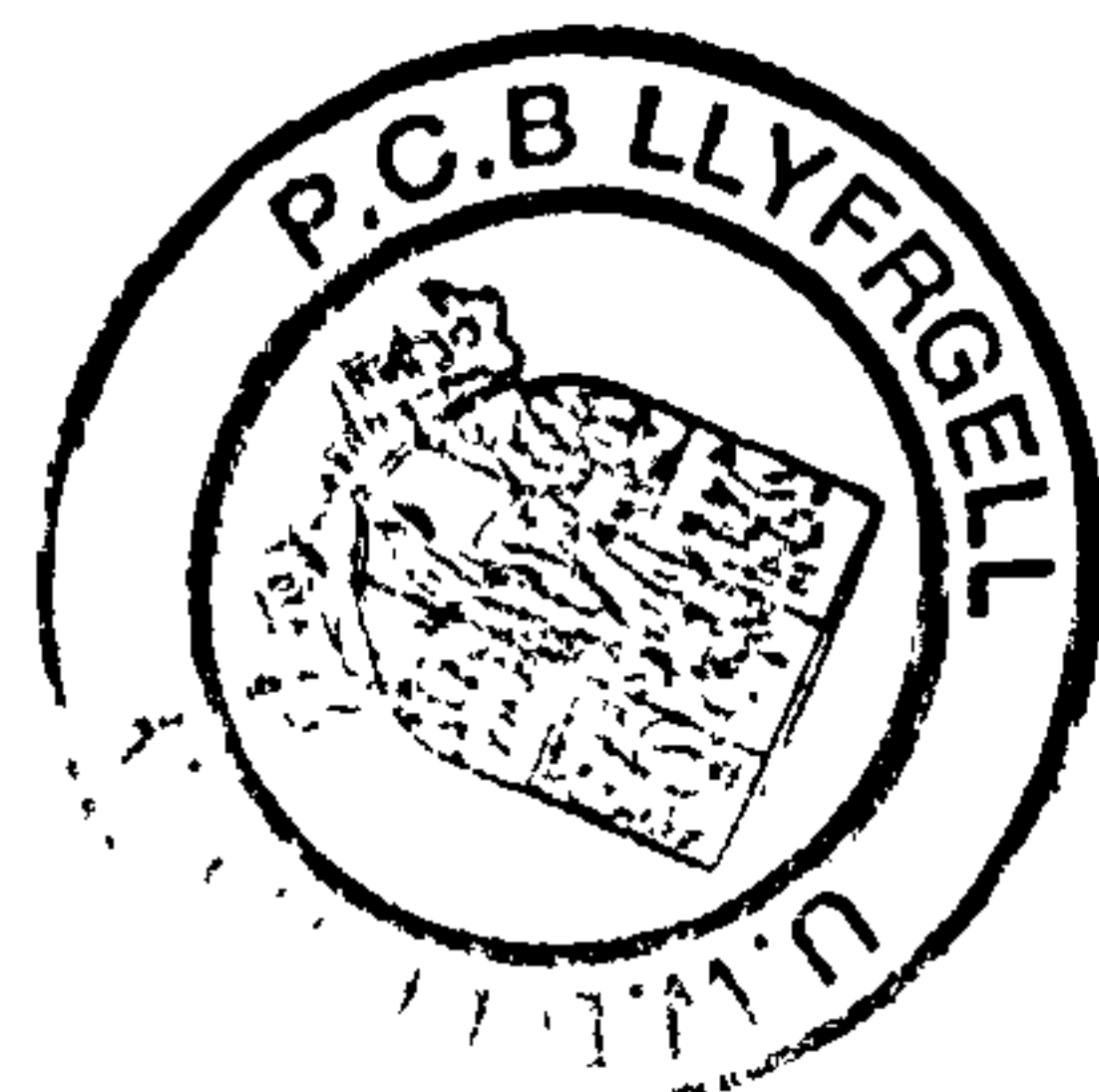
A study comparing 5 year old children's responses to a new narrative measure, with parent and teacher rated psychological problems and global concern.

**Sara Hammond-Rowley
Clinical Psychologist**



**Submitted in part fulfilment of the degree of Doctor of Clinical
Psychology, University of Wales, Bangor.**

March 1999



BEST COPY

AVAILABLE

Variable print quality

Abstract

This study explored the reliability and discriminant validity of a new narrative measure, the Teddy Bear's Picnic (Mueller, 1996), in a small non-clinical community sample of 5 year olds (n=35). The measure is designed for use with young children, and reveals internally represented psychosocial themes through the use of a story telling technique. The Teddy Bear's Picnic measure consists of nine incomplete story stems which are administered to children individually, using a range of age appropriate toys and props. Each story is based on an imaginative scenario involving a family of toy bears. Children are presented with the stories in a specific order, each of which poses an unresolved hypothetical conflict; following which they are asked to complete the unfinished story in response to the prompt "what happens next?" Concurrent parent and teacher measures were gathered as part of a larger ongoing study in the same community. Parents and teachers were asked to complete a brief 25 item rating scale, the Strengths and Difficulties Questionnaire (SDQ:Goodman, 1997), and one categorical question asking them to rate level of concern about the child's behaviour (based on the work of Stallard, 1995).

The study aimed to examine the associations between coded thematic material generated from the Teddy Bear's Picnic, and parent and teacher rated behaviour problems and global concern about behaviour. Associations were all in the expected direction, but not statistically significant. Inter-rater reliability was acceptable for total positive theme scores ($r=.91$, $p<.001$), and approached acceptability for negative total theme scores ($r=.58$, $p<.05$).

Problems were particularly apparent in the use of TBP composite total scores. Internal consistency of total theme scores and factor based subscale scores (derived from Mueller, 1996), revealed low alpha coefficients ($\alpha = .42 - .59$). Discriminant function analysis demonstrated that in this study, the TBP was unable to correctly classify children into parent and teacher rated groupings, based on both measures.

Further studies would benefit from larger samples than were available to Mueller (1996) and the present study. However, the measure appears promising as a method for eliciting psychologically relevant themes from young children, but findings suggest that further work is required in developing its psychometric properties.

Contents

| | Page |
|--|-------------|
| Abstract | i |
| Contents | ii |
| List of tables | vii |
| List of figures | vii |
| List of appendices | viii |
| Acknowledgments | ix |
| Declaration | x |
| 1. Introduction | |
| 1.1 Background to the study | 1 |
| 1.2 Child mental health | 1 |
| 1.2.1 What are child mental health problems? | 2 |
| 1.2.2 Prevalence of child mental health problems | 2 |
| 1.2.3 Influences on psychopathology | 3 |
| 1.2.4 Mechanisms of risk and resilience | 3 |
| 1.2.5 Summary | 4 |
| 1.3 Developmental processes and child mental health | 5 |
| 1.3.1 The development of emotion | 5 |
| 1.3.2 Emotion regulation | 5 |
| 1.3.3 The role of emotion in the development of knowledge | 6 |
| 1.3.4 The function of play | 6 |
| 1.3.5 Summary | 8 |
| 1.4 Measuring child mental health | 8 |
| 1.4.1 Why measure? | 8 |
| 1.4.2 Informants | 9 |
| 1.4.3 Techniques used to measure child mental health problems | 9 |
| 1.4.3.1 Psychological tests | 10 |
| 1.4.3.2 Semi-structured interviews | 10 |
| 1.4.3.3 Direct observation | 11 |
| 1.4.3.4 Questionnaires | 11 |
| 1.4.4 Summary | 12 |

| | |
|---|-----------|
| 1.5 Narrative measures | 13 |
| 1.5.1 Theoretical background | 13 |
| 1.5.2 A review of narrative measures | 14 |
| 1.5.2.1 The MacArthur story stem battery | 15 |
| 1.5.2.2 Administration of the MSSB | 16 |
| 1.5.2.3 Phases of testing | 16 |
| 1.5.2.4 Themes identified in the MSSB | 16 |
| 1.5.2.5 Coding the MSSB | 17 |
| 1.5.2.6 The Teddy Bear's Picnic | 17 |
| 1.5.2.7 Administration of the TBP | 17 |
| 1.5.2.8 Themes identified in the TBP | 18 |
| 1.5.2.9 Coding the TBP | 19 |
| 1.5.3 Studies using the MSSB | 19 |
| 1.5.3.1 Buchsbaum & Emde (1990) | 19 |
| 1.5.3.1.1 Key points | 21 |
| 1.5.3.2 Buchsbaum, Toth, Clyman, Cicchetti & Emde (1992) | 21 |
| 1.5.3.2.1 Key points | 23 |
| 1.5.3.3 Oppenheim & Emde (1996) | 23 |
| 1.5.3.3.1 Key points | 25 |
| 1.5.3.4 Kochanska, Padavich & Koenig (1996) | 25 |
| 1.5.3.4.1 Key points | 27 |
| 1.5.3.5 Oppenheim, Warren & Emde (1997) | 27 |
| 1.5.3.5.1 Key points | 29 |
| 1.5.3.6 Summary of findings from US | 29 |
| 1.5.4 Studies using the Teddy Bear's Picnic | 30 |
| 1.5.4.1 Mueller & Tingley (1990) | 30 |
| 1.5.4.1.1 Key points | 32 |
| 1.5.4.2 Mueller (1996) | 32 |
| 1.5.4.2.1 Key points | 34 |
| 1.5.4.3 Summary of findings from the UK | 35 |
| 1.5.5 Summary | 35 |
| 1.5.6 Overview of the study | 36 |
| 1.5.7 Hypotheses | 37 |

2. Method

| | | |
|------------|--------------------------------------|-----------|
| 2.1 | Design | 38 |
| 2.2 | Sample | 38 |
| | 2.2.1 Criteria for inclusion | 39 |
| | 2.2.2 Total sample meeting criteria | 39 |
| 2.3 | Measures | 39 |
| | 2.3.1 The Teddy Bear's Picnic | 39 |
| | 2.3.1.1 Reliability and validity | 41 |
| | 2.3.1.2 Factor analysis | 43 |
| | 2.3.2 Behaviour measure | 44 |
| | 2.3.3 Rating of concern | 46 |
| 2.4 | Procedure | 46 |
| | 2.4.1 Population based trial | 46 |
| | 2.4.2 TBP study | 47 |
| | 2.4.2.1 Discussions with school | 47 |
| | 2.4.2.2 Ethics approval | 47 |
| | 2.4.2.3 Piloting and training | 48 |
| | 2.4.2.4 School visits | 48 |
| | 2.4.2.5 Contacting parents | 48 |
| | 2.4.2.6 Sequence of research | 49 |
| | 2.4.2.7 Administering the TBP | 49 |
| | 2.4.2.8 Coding the TBP | 51 |
| 2.5 | Statistical analyses | 51 |
| | 2.5.1 Exploratory analyses | 52 |
| | 2.5.2 Correlation coefficients | 52 |
| | 2.5.3 Reliability analyses | 52 |
| | 2.5.4 Discriminant function analyses | 52 |

3. Results

| | | |
|------------|---|-----------|
| 3.1 | Sample characteristics | 53 |
| | 3.1.1 Non respondents | 53 |
| 3.2 | Parent and teacher ratings | 53 |
| | 3.2.1 Parent SDQ ratings | 53 |
| | 3.2.2 Teacher SDQ ratings | 53 |
| | 3.2.3 Parent ratings of concern | 54 |
| | 3.2.4 Teacher ratings of concern | 54 |
| | 3.2.5 SDQ total difficulties scores and parent ratings of concern | 55 |
| | 3.2.6 SDQ total difficulties scores and teacher ratings of concern | 55 |

| | | |
|--------------|---|-----------|
| 3.3 | TBP Narrative scores | 55 |
| 3.3.1 | Association between the measures | 56 |
| 3.3.1.1 | TBP total scores and parent SDQ scores | 56 |
| 3.3.1.2 | TBP total scores and teacher SDQ scores | 57 |
| 3.3.1.3 | Comparison of mean TBP scores across parent SDQ categories | 57 |
| 3.3.1.4 | Comparison of mean TBP scores across teacher SDQ categories | 57 |
| 3.3.1.5 | TBP total scores and parent ratings of concern | 57 |
| 3.3.1.6 | TBP total scores and teacher ratings of concern | 58 |
| 3.3.2 | Reliability of the TBP | 58 |
| 3.3.2.1 | Inter-rater reliability | 58 |
| 3.3.2.2 | Internal consistency | 58 |
| 3.3.3 | TBP factor based subscale scores | 59 |
| 3.3.3.1 | Internal consistency | 59 |
| 3.3.4 | Can TBP scores accurately predict group membership? | 60 |
| 3.3.4.1 | Classifying children who are rated as 'cases' | 60 |
| 3.3.4.2 | Classifying children who are rated as 'at risk' | 61 |
| 3.3.4.3 | Classifying children who are a cause for parent concern | 61 |
| 3.3.4.4 | Classifying children who are a cause for teacher concern | 62 |

4. Discussion

| | | |
|------------|--------------------------------|-----------|
| 4.1 | Overview of the section | 63 |
| 4.2 | Summary of findings | 63 |
| 4.2.1 | Hypothesis testing | 64 |
| 4.3 | Methodological issues | 64 |
| 4.3.1 | Sample size | 65 |
| 4.3.2 | Administering the TBP | 66 |
| 4.3.2.1 | Preparation for testing | 66 |
| 4.3.2.2 | 'Playful scenario' | 66 |

| | |
|--|------------|
| 4.3.2.3 Prompts for feelings | 67 |
| 4.3.2.4 Remaining within the 'fantasy frame' | 68 |
| 4.3.2.5 Video-resording the TBP | 68 |
| 4.3.3 Scoring the TBP: | |
| Interpreting the scoring directions | 68 |
| 4.3.3.1 General scoring rules | 68 |
| 4.3.3.2 Using specific themes | 70 |
| 4.3.4 The use of composite total scores | 73 |
| 4.3.5 Factor analysis | 73 |
| 4.3.6 Improving reliability | 75 |
| 4.4 Clinical findings from the TBP | 75 |
| 4.4.1 Positive themes | 75 |
| 4.4.2 Negative themes | 76 |
| 4.4.3 Sex differences | 77 |
| 4.4.4 Statistical vs clinical significance | 78 |
| 4.5 Integration of findings | 78 |
| 4.5.1 Comparison of TBP and MSSB | 79 |
| 4.5.2 Findings from previous studies | 80 |
| 4.6 Implications for the measurement of child mental health | 83 |
| 4.7 Theoretical implications | 84 |
| 4.8 Therapeutic implications | 85 |
| 4.9 What next? Further research using the TBP | 86 |
| 4.10 Conclusion | 88 |
| 5. References | 90 |
| 6. Appendices | 108 |

List of Tables

| | Page |
|---|-------------|
| Table 1: Themes sought from each TBP story stem | 42 |
| Table 2: Threshold scores for parent and teacher SDQ measure (Goodman, 1997) | 45 |
| Table 3: Sequence of research activity | 50 |
| Table 4: Distribution of scaled scores for parent and teacher completed SDQ's (n=35) | 54 |
| Table 5: Distribution of parent and teacher completed ratings of concern | 55 |
| Table 6: Distribution of TBP total scores by sex (all items) (n=35) | 56 |

List of Figures

| | |
|-----------------------------------|----|
| Figure 1: TBP study sample | 40 |
|-----------------------------------|----|

List of Appendices

| | | Page |
|---------------------|--|-------------|
| Appendix 1: | Teddy Bear's Picnic (Mueller, 1989): Complete set of story stems | 108 |
| Appendix 2: | SDQ: The strengths and difficulties questionnaire (Goodman, 1997) parent and teacher versions | 114 |
| Appendix 3a: | Concern question (parent version) | 115 |
| 3b: | Concern question (teacher version) | 116 |
| Appendix 4: | Ethics approval letters: NWAHA University of Wales, Bangor; Psychology Department | 117 118 |
| Appendix 5: | Parent consent letter (English version) (Welsh version) | 119 120 |
| Appendix 6: | Thank you letters | 121 |
| Appendix 7: | Teddy Bear's Picnic: Scoring directions (Revised version) (Mueller, 1996) | 124 |
| Appendix 8: | Teddy Bear's Picnic: Scoring form (Mueller, 1996) | 135 |

Acknowledgements

I would like to thank the school that participated in the study, especially the Headteacher and class teachers who helped enormously with the practical organisation of the research. I would particularly like to thank the children, and their parents for agreeing to take part, without whom the study would not have got off the ground.

I am very grateful to the Clwydian Community Care NHS Trust, Department of Clinical Psychology for funding the degree, and to my line managers Dr Peter Appleton and Dr Colin Elliott for their continued and generous support. I am especially grateful to Dr Peter Appleton, who patiently supervised the research, and provided guidance, encouragement and practical help throughout; despite my deadline avoidance behaviour.

Thanks are also extended particularly to Dianne Cairns, Child Mental Health Therapist, and Joanne Maher, Research Assistant for uncomplainingly collecting most of the data; to Lee Murphy for coding the video material; to Dr Mike Startup and Lisa Rowland for statistical advice; and to Dr Helen Sharp for enthusiastic Teddy Bear talk. Thanks also to Dr Robert Jones and staff at the North Wales Clinical Psychology Course for organising the workshops, and to Sue Robinson and Chris Gilbert for support and encouragement.

On a personal note, massive thanks to Mark Rowley for unremitting practical and emotional support during these recent months, and to friends and family who've got used to not hearing from me as often as usual.

(1997). Concern for the mental health of children has increased in recent years, partly as a result of growing evidence to confirm that children can experience significant psychological symptoms and disorders as a result of a variety of individual and psychosocial factors; and also in response to the increasingly recognised associations between early emotional and behavioural problems and later adolescent and adult psychopathology, and criminality (Smith, 1995; NHS/HAS, 1995).

1.2.1 What are child mental health problems?

Child mental health problems can be defined as difficulties involving emotions or behaviours (or both) which interfere with a child's development, learning or social relationships (Achenbach, 1995). Types of problems include both 'internalising' and 'externalising' problems e.g. withdrawn behaviour, anxiety and depression, somatic complaints; and aggression, disruptive behaviour, attention problems, poor social skills, and peer rejection. These two broad clusters of childhood symptoms essentially represent groups of behaviours which demonstrate behavioural deficits or inhibitions (over controlled 'internalisers'), and behavioural excesses (under controlled 'externalisers') (Achenbach & Edelbrock, 1978). Across cultures, under control problems are found more often in boys, and over control in girls (Weisz, Suwanlert, Chailiyasit & Walter, 1987).

1.2.2 Prevalence of child mental health problems

Serious emotional and behavioural problems are very common in school aged children. A number of studies demonstrate a child mental health prevalence of between 13% (Institute of Medicine, 1994) and 20% (Shepherd, Oppenheim & Mitchell, 1991; Verhulst & Koot, 1995). Knowledge of the prevalence of mental health problems in the pre and early school phase is less well developed. One study of the prevalence of emotional and behavioural problems in a total population of 3 year olds revealed a 7% rate of moderate to severe behaviour problems, and a further 15% with mild problems (Richman, Stevenson & Graham, 1982). Factors associated with the development of problems in this study included family influences such as marital relationship difficulties, maternal mental state, and external stresses; and individual child factors such as language development and temperament. This study also demonstrated that 60% of children identified as having mental health problems at the age of 3 years were still experiencing difficulties five years later. This suggests that it is in the very early years of life that many problems first begin,

1. Introduction

1.1 Background to the study

Whilst child mental health problems are not new phenomena, recent advances in epidemiological research in clinical child psychology and psychiatry have raised important questions about the development of childhood disorders, and issues relating to the early identification and prevention of child mental health problems. The questions on which this study is based began to germinate during a community child mental health needs assessment, in preparation for the development of a new service, the principle aims of which are related to problem prevention in children. Current research evidence supports the premise that early intervention optimises treatment effectiveness, yet in identifying which children would benefit from intervention, the children themselves rarely act as their own informants. Instead, they are brought to the attention of services by adults (usually parents or teachers) at a stage when difficulties have already begun to interfere with development, and often when the child's behaviour has become significantly troublesome to those around them. This study seeks to make a contribution to the development of an appropriate measure for direct use with young children, and in so doing; increasing the opportunity for the 'child's voice' to be heard at an early stage.

The introduction provides an overview of theoretical and research material relevant to the study, which was conducted with a small community sample of 4-5 year old (reception age) children attending one local primary school. The section starts by defining child mental health problems, influences on psychopathology and mechanisms of risk and resilience; before moving on to discuss developmental processes and their central role in the onset and maintenance of emotional and behavioural difficulty. Measurement issues are summarised, prior to a detailed review of the theoretical and research background of narrative techniques, and their use to date with young children. The section concludes with a review of seven recent studies which have used narrative measures in eliciting information directly from young children.

1.2 Child Mental Health

The health and development of children has been of high public and private concern for decades, and is reflected in a range of reports over the last forty years e.g. Ministry of Education (1955); Caplan (1964); World Health Organisation (1972, 1982); Home Office (1991); Kolvin, Miller, Fleeting & Kolvin, (1988); Welsh Office

(1997). Concern for the mental health of children has increased in recent years, partly as a result of growing evidence to confirm that children can experience significant psychological symptoms and disorders as a result of a variety of individual and psychosocial factors; and also in response to the increasingly recognised associations between early emotional and behavioural problems and later adolescent and adult psychopathology, and criminality (Smith, 1995; NHS/HAS, 1995).

1.2.1 What are child mental health problems?

Child mental health problems can be defined as difficulties involving emotions or behaviours (or both) which interfere with a child's development, learning or social relationships (Achenbach, 1995). Types of problems include both 'internalising' and 'externalising' problems e.g. withdrawn behaviour, anxiety and depression, somatic complaints; and aggression, disruptive behaviour, attention problems, poor social skills, and peer rejection. These two broad clusters of childhood symptoms essentially represent groups of behaviours which demonstrate behavioural deficits or inhibitions (over controlled 'internalisers'), and behavioural excesses (under controlled 'externalisers') (Achenbach & Edelbrock, 1978). Across cultures, under control problems are found more often in boys, and over control in girls (Weisz, Suwanlert, Chailiyasit & Walter, 1987).

1.2.2 Prevalence of child mental health problems

Serious emotional and behavioural problems are very common in school aged children. A number of studies demonstrate a child mental health prevalence of between 13% (Institute of Medicine, 1994) and 20% (Shepherd, Oppenheim & Mitchell, 1991; Verhulst & Koot, 1995). Knowledge of the prevalence of mental health problems in the pre and early school phase is less well developed. One study of the prevalence of emotional and behavioural problems in a total population of 3 year olds revealed a 7% rate of moderate to severe behaviour problems, and a further 15% with mild problems (Richman, Stevenson & Graham, 1982). Factors associated with the development of problems in this study included family influences such as marital relationship difficulties, maternal mental state, and external stresses; and individual child factors such as language development and temperament. This study also demonstrated that 60% of children identified as having mental health problems at the age of 3 years were still experiencing difficulties five years later. This suggests that it is in the very early years of life that many problems first begin,

and that problem continuity is high.

1.2.3 Influences on psychopathology

A variety of complex factors and mechanisms interact and contribute to the development of individual child psychopathology. These include genetic influences (Simonoff, McGuffin & Gottesman, 1994); chromosomal abnormalities (Bolton & Holland, 1994); organic brain disorders (Goodman, 1994a); chronic adversities (Garnezy & Masten, 1994); physical and emotional maltreatment (Skuse & Bentovim, 1994); sexual abuse (Smith & Bentovim, 1994); chronic physical illness (Taylor & Eminson, 1994); and disrupted caregiving (Wolkind & Rushton, 1994; Hersov, 1994). In addition to individual factors within the child, environmental factors in the form of social and economic influences also play a significant part in contributing to the onset and development of disorder (Institute of Medicine, 1994). It is particularly important to study the distribution of disorders within whole populations if the incidence (the rate at which new cases of a disorder arise) and prevalence (the proportion of the population with each disorder) of child mental health problems are to be reduced through preventive and treatment interventions (Rutter & Smith, 1995).

1.2.4 Mechanisms of risk and resilience

Factors which predispose children to an increased probability of developing mental health difficulties are generally referred to as risk factors. These can exist within the individual child, their family, or the community; and can be biological or psychosocial in their origin (Werner & Smith, 1992). A number of research studies suggest that in general, a single isolated stressful or adverse experience does not pose significant risk for the development of mental health difficulties in children (e.g. Emery, 1982; Kolvin, Charles, Nicholson, Fleeting & Fundundis, 1990; Rutter & Smith, 1995). Serious risk tends to occur when a combination of stresses occur simultaneously, or when stresses or adversities accumulate over time, leading to direct and indirect cumulative mechanisms acting in tandem to increase the likelihood of further negative experiences and psychological distress (Rutter, 1989). Some risk factors are specific to a particular disorder, but most are common to many disorders. For example, factors which increase children's vulnerability to conduct disorder include biological risk factors in the form of genetic vulnerability, physiological abnormality, gender, early temperamental problems, hyperactivity, cognitive and neuropsychological difficulties and chronic illness. Psychosocial risk factors include

aggressive behaviour, peer rejection, maternal depression, marital discord, family violence, substance misuse, criminality, and harsh and punitive parenting style.

Community risk factors include socioeconomic disadvantage, poor housing, large family structure, urban residence, and social disintegration (Rutter & Smith, 1995).

Children vary in their ability to be resilient to psychosocial adversity (Rutter, 1987). This involves an interaction between individual vulnerability, risk factors and protective mechanisms. Rutter (1985) defined protective factors as “those factors that modify, ameliorate or alter a person’s response to some environmental hazard that predisposes to a maladaptive outcome”. Resilience has previously been viewed as an individual characteristic, and protective factors as the consequence of positive experience (Rutter & Smith, 1995). This view has been modified in light of evidence which suggests that resilience is derived from compensatory experiences such as a positive relationship with one parent (Jenkins & Smith, 1990), previously secure separation experiences (Stacey, Dearden, Pill & Robinson, 1970), or successful coping with psychosocial stress (Rutter, 1993). For example, factors which reduce the probability that a child will develop conduct disorder include good intelligence, academic achievement, an ability to get along with parents, siblings, peers and teachers, prosocial behaviour, and involvement in non-school activities (Rutter, 1985).

1.2.5 Summary

Child mental health problems can be described as difficulties of an emotional or behavioural nature, which tend to cluster into internalising or externalising patterns of behaviour. Epidemiological studies suggest a prevalence rate of child mental health problems as between 13% and 20%, and concern has increased as several associations between early childhood difficulties and adolescent and adult psychopathology have been confirmed through longitudinal research. Factors which contribute to the development of psychopathology include individual characteristics and traits, and mechanisms of risk and protection. Some risk factors are specific to one disorder, but most are common to several, suggesting that a risk reduction approach to interventions may significantly reduce the prevalence of child mental health problems.

1.3 Developmental processes and child mental health

1.3.1 The development of emotion

In the context of child mental health, emotional processes are critical to increasing understanding of how mechanisms of risk and protection operate within the individual child. Children's capacity to regulate emotional processes becomes relevant in understanding the development of problems, and in differences between individuals in their ability to modulate arousal and its expression.

1.3.2 Emotion regulation

The progress of emotional development is intertwined with changes in social, moral, cognitive and language development. In its broadest sense, emotional development involves a movement from dyadic regulation (within the context of the caregiving relationship) to self regulation (within the context of autonomy and independence) (Sroufe, 1989). Bowlby's work on the function and processes of attachment relationships provides considerable insight into the importance of early attachment patterns and the dyadic regulation of emotion, and later adaptation and self regulation of emotion (Bowlby, 1973). The influence of early attachment on later adaptation is not linear in a causal sense, but the association between infant attachment representations and later disorder supports the view that this process is central in influencing the development of emotional regulation through the very early guidance and support of the caregiver. Bretherton (1990) suggests that emotionally open communication between caregivers and children assists young children in their ability to form internal 'working models' of their attachment relationships which are understandable and coherent.

Allen, Moore, Kuperminc & Bell (1998) found an association between attachment organisation and psychosocial functioning in adolescents. Secure attachment experiences were related to competence with peers, and lower levels of internalising and deviant behaviours, in contrast to adolescents with insecure attachment experiences. The study suggests that security is linked to later psychosocial functioning via mechanisms which are responsible for the internal organisation of affect and cognition. Several earlier studies have also demonstrated strong associations between early disruption in attachment relationships and later disruption in emotion regulation (Collins & Read, 1990; Patrick, 1994).

Children begin to engage with their environment from birth. During the pre school

years, securely attached infants are more able to contain affect, impulsivity and behaviour appropriately, whilst at the same time being able to flexibly express feelings through the activities in which they are engaged e.g. learning, play and social relationships. In contrast, emotional dysregulation is manifested in a variety of affect and behaviour indicative of a less secure attachment history e.g. tense, anxious, easily frustrated, passive, hostile, aggressive antisocial or aloof (Schore, 1994; Sroufe, Carlson & Shulman, 1993; Urban, Carlson, Egeland & Sroufe, 1991; Maccoby, 1980).

1.3.3 The role of emotion in the development of knowledge

Emotional development and regulation are influential in the acquisition of knowledge, skill and consciousness. The constructs of procedural and declarative knowledge are central to the development of consciousness during early childhood (Emde, Biringen, Clyman & Oppenheim, 1991). Young children demonstrate procedural knowledge in their interactions with caregivers from infancy onwards (Lewis & Brooks-Gunn, 1979). This takes the form of skill acquisition based on information which is not necessarily available at a conscious level (Cohen & Squire, 1980). Declarative knowledge develops at a later stage and is based on information which can be brought into conscious awareness, after the stage at which infants develop the capacity for reflective self awareness. Emotionally engaging experiences from infancy are therefore stored procedurally and cannot usually be accessed at a conscious level (Emde et al., 1991), which has important implications for assessing children's internal representations of themselves and others. Knowledge, skill and emotion all develop in tandem, and a process which links all three is found in children's play.

1.3.4 The function of play

A central process which contributes significantly to the increase in emotional regulation in children is the development of fantasy play. Sroufe (1996) suggests that the capacity for play is not sufficiently appreciated in understanding children's emotional development. During the preschool period, new abilities are reflected in new forms of play (Belsky & Most, 1982; Bretherton & Bates, 1985). As children reach the end of infancy (18-24 months), symbolic play begins to occur which can be observed in children's ability to use one object to represent another (e.g. a stick to represent a spoon). Smith (1990) describes four main types of play which provide practice situations for later functions and adaptive behaviours. These are locomotor

play (e.g. running, jumping), object play (e.g. learning to handle things and manipulate them), social play (e.g. physical 'rough and tumble', and non physical games), and fantasy play (i.e. where the meanings of objects and actions are transformed to fit an imaginary situation).

The development of early play and pretence activity has been shown to occur at similar ages in all normally developing children (Farver, 1992). It coincides with early language development (McCune, 1995), and follows a similar course, reaching a peak in the preschool years (Singer & Singer, 1990). Play has been found to provide a similar adaptive function in different circumstances and cultures (Weinberger & Starkey, 1994; Tamis-LeMonda, 1992). The assessment of play, and play therapy within clinical settings has led to rich information about children's individual experience, and several assumptions have been made about the meaning associated with particular forms of pretend play (Mayes & Cohen, 1993). However, empirical research has been patchy and inconclusive in determining how useful or accurate these assumptions are in understanding the child's own perspective of his or her own experiences (Cohen, Velez, Kohn, Schwab-Stone & Johnson, 1987).

By the time children begin school, they have achieved several developmental milestones, and have experience of three years of verbal interaction with their environment. Skills are being acquired in the domains of cognition, language, and social interaction, and moral values are forming and beginning to be assimilated into everyday activities. From the age of 3 years, children begin to use narrative expression within play to make sense of their experience and share it with others. Children practice and increase their understanding of interaction patterns through engaging in fantasy play, and the world of 'imaginative reality' and 'imaginative pretence' (Emde, Kubicek, Oppenheim & Oppenheim, 1997). Several authors suggest that children are not confused by these two forms of experiential play, but rather make use of their differences to reflect on aspects of emotional regulation, and manage feelings which have been generated as a result of everyday experience. The use of narratives within this context allows the child to practice alternative possibilities and outcomes, and forms an early mechanism of emotional problem solving and regulation. This ability reflects a major development in play and language which occur between the ages of 2 and 3 years.

1.3.5 Summary

The process of emotion regulation is of particular relevance to child mental health, and understanding normative processes and individual differences in tandem increases knowledge of risk and resilience in response to adversity. During infancy, children's representations of emotional experiences take the form of procedural knowledge, which is not accessible at a conscious level. As developmental milestones occur, and language is acquired, declarative knowledge becomes available at a conscious level, and children begin to experiment with play and narrative expression to test out adaptive responses to moral and emotional dilemmas.

1.4 Measuring child mental health problems

1.4.1 Why measure?

Measuring child mental health problems has three main purposes, and can be carried out at three levels. Firstly, children present to clinical settings for help with a variety of emotional and behavioural problems. These require definition by the clinician to increase understanding of the difficulty, and measuring specific aspects of the presenting problem contributes to the overall process of information collection and assessment. The objective of measurement in this context relates to problem formulation, and is carried out on an individual basis.

A second purpose of measuring child mental health problems relates to therapy outcome. A range of characteristics and symptoms are common across a variety of presenting problems. These can be measured using a variety of techniques, and if administered before and after therapy, provide insight into the effectiveness of the steps taken in providing a particular intervention. The objective of measurement in this context relates to evaluating the impact of an intervention, and can be carried out on an individual or group basis.

A third purpose of measuring child mental health problems has its foundation in epidemiology. The objective of measurement in this context is to assess the distribution of particular disorders within populations; to identify processes and mechanisms associated with the presence or absence of disorder; to identify individuals or groups of individuals who are most at risk for developing disorder; and to evaluate the impact of interventions in aiming to prevent onset. Knowledge of these factors also contributes to the process of service planning and resource

distribution e.g. in developing child and adolescent mental health services across geographical areas.

1.4.2 Informants

Several methods for assessing young children's mental health problems rely entirely on information gathered from adult informants, predominantly parents and teachers (Berger, 1994). Additional sources of information include official agencies such as Social Services, children themselves, and their peers (Scott, 1996). Offord, Boyle, Racine, Szatmari, Fleming, Sanford & Lipman (1996) provide evidence to demonstrate that child psychiatric disorders should be conceptualised as informant and setting specific. This is because adults see children in different settings (e.g. home and school) and children frequently behave differently in each context. Several studies have shown that adults in the same setting also differ in their perceptions of the same behaviour (Scott, 1996) e.g. mothers tend to report more misbehaviour than fathers; agreement between parents, and parents and teachers increases with increasing problem severity.

Self report measures are well established with older children and adolescents e.g. Youth Self Report Questionnaire (Achenbach & Edelbrock, 1987), the Children's Depression Inventory (Kovacs, 1982). However, several studies have shown that the collection of self report information from young children is problematic, lacking reliability and validity in children younger than 8 years (Perrin & Last, 1992). Clinicians have also noted the tendency for young children to under report problems in clinical interview situations (Glasberg & Aboud, 1982). Gathering information from children's peers in peer nomination tasks provides highly relevant sociometric data (Coie & Dodge, 1988), but does not seek information directly from children about themselves. However, the systematic understanding of children's mental health problems cannot be complete without some inclusion of the child's own perspective on his or her difficulty (La Greca & Silverman, 1993), yet the development of appropriate direct assessment methods which are reliable and valid has proved difficult. Recent studies demonstrate some progress in refining techniques (Sturgess & Ziviani, 1995; Herjanic & Reich, 1997), and several authors comment on the indispensability of the child or adolescent acting as their own informant, in addition to parent and teacher information (Verhulst & Ende, 1992a). This supports the view that research into developing techniques to gather information directly from young children is a priority.

1.4.3 Techniques used to measure child mental health problems

Techniques used in the measurement of young children's mental health problems can be summarised into four main categories consisting of psychological tests, semi-structured interviews, observation systems, and questionnaires.

1.4.3.1 Psychological tests

Berger (1994) describes psychological tests as "procedures for measuring individual differences in psychological characteristics". They can be selected to focus on specific characteristics of interest e.g. intellectual ability, spatial reasoning, memory; where several items are used to sample the same characteristic in a standardised format. This reduces the likelihood of subjectivity in the testing procedure, and of chance error in performance. Standardised psychological tests are explicit in their purpose, replicable, and allow comparisons to be made against data from large samples, which enables conclusions to be drawn about the extent of deviation from the norm. This has potential value in identifying specific disorder (Garfield, 1974). Examples of psychological tests used with children include the Wechsler Intelligence Scale for Children (Wechsler, 1974), and the Rivermead Behavioural Memory Test (Wilson, Ivani-Chalian & Aldrich, 1991).

Projective psychological tests are designed to elicit a wide range of complex responses which are believed to be reflective of internally represented underlying psychological processes in the respondent (Gleser, 1963). Chazan (1970) describes the usefulness of projective techniques with children in clinical settings in stimulating the imagination of the child through the creation of a make believe situation. The child is encouraged to speak or act out responses to stimuli as the test proceeds, the aim of which is to uncover conscious and unconscious drives, feelings, cognitions and behaviours.

Projective techniques are usually administered according to a standard procedure, but lack objectivity in their interpretation (Griffiths, 1970). Examples of projective psychological tests designed for use with children include the MAPS (Make a Picture Story) technique (Schneidman, 1947); the Raven Controlled Projection Test (Raven, 1951); the Sentence Completion Test (Rotter & Rafferty, 1950); and the Draw - a - Person and Kinetic Family Drawing (Tharinger & Stark, 1990).

1.4.3.2 Semi - structured interviews

Scott (1996) summarises the salient features of semi-structured interviews in the measurement of child mental health problems. Semi-structured interviews consist of a series of detailed questions presented in an exact form, which increases validity in administration, but allows the interviewer to probe further than the first response when allocating ratings. Taylor, Schacher, Thorley & Weislerberg (1986) state that semi-structured interviews are more closely correlated with other measures of antisocial behaviour in children, and are more accurate in predicting long term outcome when compared with questionnaires and rating scales. Drawbacks include the potential for informants to be imprecise about specific frequencies, antecedents and consequences of behaviours, but in contrast, informants are generally more accurate in the reporting of discrete events. Examples of semi-structured interviews devised in Britain for use in child mental health include the Child and Adolescent Psychiatric Assessment (CAPA: Angold, Prendergast, Cox, Harrington, Simonoff & Rutter, 1995). This interview covers all aspects of psychopathology and impairment, and includes a parent and child version. The Parent Account of Child Symptoms (PACS: Taylor, Schacher, Thorley & Weislerberg, 1986) is a parent interview specifically aimed at eliciting information about conduct problems and hyperactivity, with comprehensive concurrent and predictive validity data.

1.4.3.3 Direct Observation

Observation methods focus on the moment to moment process of interaction between the child and his or her environment as well as the specific nature of the behaviour in question (Scott, 1996). Direct observation has several advantages over informant based methods of measurement e.g. it does not rely on a third person description of behaviours and events. Drawbacks of observation include a lack of applicability to emotional difficulties, or behaviours which are covert; limitations imposed by the observer which may influence 'real' behaviour; and the inevitably limited sample of behaviour that can be observed at any one time. Examples of British observation schemes include the Dowdney, Mrazek, Quinton & Rutter system (1984) which is designed to assess general parenting behaviour in families of 2-3 year old children; and the Gardner (1994) system which assesses family interactions in oppositional and aggressive children.

1.4.3.4 Questionnaires

Questionnaires and rating scales consist of a series of statements or questions

which respondents are asked to complete according to brief instructions. They are generally easy to use and quick to complete, usually by selecting one of a limited choice of possible responses. Questionnaires are particularly useful in screening for child mental health problems, as they have high face validity, and provide a selection of response categories which do not require lengthy consideration for completion. This increases the likelihood of higher response rates in population based studies, where the purpose of administration is to establish a broad measure of prevalence, severity and comorbidity. Drawbacks of questionnaires include the increased influence of rater bias, and the lack of individualised detailed information which reduces the instrument's sensitivity to small changes in target problems. Examples of questionnaires developed to measure child mental health problems include the Achenbach Checklists (Achenbach, 1991) which consist of parent and teacher versions, and a self report version for older children (11-18 years); and the Conners Rating Scales (Goyette, Conners & Ulrich, 1978) which are designed to measure hyperactive behaviour, and are available in parent and teacher versions, both in short (CPRS-48, CTRS-28) and long forms (CPRS-93, CTRS-39).

1.4.4 Summary

Child mental health problems can be measured on an individual, group or population level. Measures are useful in contributing to our knowledge and understanding of presenting problems in clinical settings; in studies of intervention effectiveness; and in identifying the distribution and prevalence of disorder within communities. Information about young children's mental health problems is most often gathered from adult informants, and has been shown to be informant and setting specific. Additional measures of child mental health problems can be gathered from peers and agency data.

There is a demonstrable lack of reliable and valid methods for obtaining information from children themselves in the form of direct measurement techniques. Techniques most often used in measuring child mental health problems include psychological tests, semi-structured interviews, direct observation and questionnaires. The measurement of child mental health problems is never an end in itself, but rather a means to discovering meaningful information about predisposing, precipitating and maintaining factors, and mechanisms of risk and resilience. The lack of information gathered directly from children highlights a significant gap in our understanding of these processes from the child's perspective.

Increasing knowledge of early development in the infancy and preschool stage, together with the recognised lack of direct methods for assessing young children, has led to the exploration of the usefulness of direct techniques which focus on children's play and narrative expression in understanding child mental health.

1.5 Narrative Measures

1.5.1 Theoretical Background

Research has suggested that children are able to distinguish between pretence and reality at a very early age, and that they make use of this ability in regulating emotion during everyday interaction (Emde et al., 1997, Golumb & Balasso, 1995). Studies which have led to this conclusion have taken a variety of forms, including controlled observation (Flavell, 1986; Harris & Kavanaugh, 1993), naturalistic observation (Wolf, 1984), and observation of play with siblings in home settings (Dunn & Dale, 1984); and are nested in a variety of theoretical perspectives (e.g. cognitive-experimental, psycholinguistic, social learning, attachment, and psychoanalytical).

Between the ages of 2 and 3 years, children's expressive language develops from one word to multiple word speech (McCune, 1995), alongside the development of self awareness and the ability to be empathic (Eisenberg & Miller, 1987; Hoffman, 1983). At this stage, children also begin to engage in narrative speech, which enables them to make sense of their experience and share it with others (Emde et al., 1997). Children begin to develop the capacity for narrative speech between the ages of 2 and 3 years, and this takes the form of the verbal expression of feelings and wishes in the form of a story. For example, at 2 years, narrative expression will take the form of one or two word responses prompted by a caregiver (Emde et al., 1997). At 3 to 4 years, children are able to respond imaginatively to prompts using sentences containing 8 or more words, engaging in a process of playful interaction during which they can switch between pretence and reality with ease. This allows children to begin to explore and test out solutions to dilemmas in an imaginary world, in advance of trying out such solutions in reality, thus providing internal feedback in preparation for coping with real situations (Bruner, 1986). Emde et al., (1997) describe this as the capacity for 'two kinds of psychic reality', namely 'imaginative pretence' (fantasy) and 'imaginative reality' (a representation of everyday reality involving imagination); allowing the child to develop early problem solving skills through their ability to engage in playful activities. Early speech, and particularly narrative expression, is central to this process of distinguishing reality from

imagination.

Several studies have highlighted the importance of coherent narrative organisation in early mother-child conversation and the development of emotionally coherent narratives in children (Hudson, 1990; Fivish, 1991). Associations have also been established between the ability to construct emotionally coherent narratives and emotion regulation (Oppenheim, Nir, Warren & Emde, 1997).

1.5.2 A Review of Narrative Measures

The development of narrative techniques began during the early 1990's in response to findings from developmental research, which consistently supported the view that children are able to distinguish between pretence and reality in their play (e.g. Wolf & Slade, 1994). In recognising that children make use of narrative expression in spontaneous play, methods for the direct assessment of thoughts, feelings and experience through the use of stories and pretend activity have begun to emerge.

Narrative measures may usefully be compared with projective psychological tests in two main ways. Firstly, in projective tests, children's spontaneous responses to indirect stimuli (e.g. pictures, incomplete sentences) generate information which contributes to the process of formulating the child's psychological presentation. Similarly, in narrative measurement, incomplete story stems act as the indirect stimuli, and the verbal and non verbal responses from the child form an information base from which conclusions can be drawn. Secondly, the meaning which is attached to the responses given in both projective psychological tests and narrative measures is based on evidence which supports the premise that internal representations can be tapped through indirect, age appropriate activities e.g. telling stories through toy/doll play, drawing pictures.

Whilst there are similarities between projective tests and narrative measures, there are several differences which are apparent and substantial. These relate specifically to the theoretical base from which the two methods have developed, and the process of attaching meaning to the material generated in response to the specific stimuli. Projective psychological tests are rooted in psychoanalytic theory, in contrast to narrative measures which embrace a range of theories including attachment (Bowlby, 1988), social learning (Bee, 1992), narrative theory (Murray, 1995), and psychoanalytic theory (Emde et al., 1997). In projective testing, material is

interpreted in a largely subjective manner, and relies entirely on the individual skills and perspective of the tester. Projective test material can offer considerable insight into individuals in clinical settings, but contributes minimally to the understanding of children in groups or whole populations; as findings are uncontrolled and lack standardisation (Griffiths, 1970). In contrast, the material generated from narrative measures is summarised through a standardised scoring procedure, which minimises subjectivity through coding only what is directly observable within the testing environment. Conclusions which are drawn about the responses given focus on the child's current presentation, in contrast to projective test material, which makes assumptions about the development of specific responses in the context of psychoanalytic theory.

Two measures have been developed and tested, The MacArthur Story-Stem Battery (Bretherton, 1990b), and The Teddy Bear's Picnic (Mueller and Tingley, 1990, Mueller, 1996). These measures are reviewed below.

1.5.2.1 The MacArthur Story-Stem Battery

The MacArthur Story-Stem Battery (MSSB) consists of a standardised set of story beginnings (stems) which can be administered to children between the ages of three and seven years. The complete version consists of in excess of twelve stems (Warren, Emde & Oppenheim, 1993), where specific stories are chosen to explore specific themes and representations. The MSSB was developed in the United States, initially by Warren, Mantz-Simmons & Emde (1993), as part of the MacArthur Research Network for Early Childhood Transitions. Several authors have added stems to explore additional emotional themes within the context of their particular research (e.g. Kochanska, DeVet, Goldman, Murray & Putman, 1994). The stories are designed to develop an understanding of the child's perception of family relationships, conflicts and emotional situations, using a developmentally appropriate technique (Bretherton, 1990b; Bretherton, Ridgeway & Cassidy, 1990; Buchsbaum & Emde, 1990; Buchsbaum, Toth, Clyman, Cicchetti & Emde, 1992). The battery can be administered in its entirety, or specific stories can be selected to explore specific themes e.g. aggression. The story stems in the MSSB cover a variety of events ranging from childhood transgressions, separation issues and loss, to situations leading to moral dilemma e.g. 'monster in the dark', 'the bathroom shelf', 'exclusion', 'reunion'.

1.5.2.2 Administration of the MSSB

Each stem is delivered in a structured standardised form, then children are prompted to finish the story by being asked “show me and tell me what happens now”.

Prompts within each story stem are specific and clearly defined, and any additional information presented by the examiner may render the story, or parts of the story invalid. Props are used throughout in order to engage children in the story-telling process. These take the form of developmentally appropriate toys and dolls.

Administration of eleven out of twelve story stems took an average of forty minutes in one study with children aged four years six months and five years six months (Oppenheim et al., 1997).

1.5.2.3 Phases of Testing

Responses to each story are divided into three phases. The presentation phase focuses on the child’s response to the initial presentation of the toys and props at the beginning of each story. The narrative development phase begins from the point at which the presenter asks the child “show me and tell me what happens now”. The transition between narratives phase begins when the story is clearly finished, but prior to the delivery of the subsequent stem. Total time for each phase is recorded.

1.5.2.4 Themes identified in the MSSB

Themes from the battery are separated into three groups. Content themes can be present throughout all phases of each story, and consist of developmentally salient relationship themes (e.g. competition, sharing, rivalry, exclusion, aggression, empathy, interpersonal conflict resolution, ‘tattling’); moral issues (e.g. compliance, non-compliance, escalation of conflict, shame, affiliation, blame, affection, dishonesty, punishment, reparation, verbal conflict, personal injury); atypical responses (positive, neutral and/or negative); and repetition (e.g. of the story line as just presented, or recurring from a previous stem). Additional content themes which can be coded include danger, disappointment/ loss, destruction of objects, oral expression, difficulty telling the narratives, story telling technique, knowledge of emotions, existence of the monster (child’s belief), and parental view of the monster (parental response).

Parental representation themes include positive actions or statements (e.g. protective, care taking, affectionate/warm/caring, helpful); negative actions or statements (e.g. harsh/punitive, rejecting, ineffectual); discipline (e.g.

discipline/control which may include physical punishment which is not too harsh); and 'parents as people' (e.g. when the child represents feelings from a parental perspective).

Performance themes relate to the child's overall performance throughout all phases of the procedure, and include control, emotional openness (e.g. personal expressivity of joy, anger, distress, concern, sadness, anxiety, denial); role of parents; narrative coherence; indirect vs direct performance style, relationship with the examiner (e.g. responsiveness with examiner, involvement of examiner), investment in performance, indicators of the child's understanding of the conflict; final affect (e.g. positive, neutral, negative, fear, angry); and positive to negative emotional incoherence.

1.5.2.5 Coding the MSSB

The MSSB is videotaped for later coding. Instructions relating to coding categories are provided in a detailed manual. Information is gathered from the child's verbal and non verbal responses to the presentation, narrative development and transition phases of each story. The start and end time for each phase is noted, and total time for each phase is calculated at the end of coding. Comments are noted throughout, including those which relate to examiner error.

1.5.2.6 The Teddy Bear's Picnic

The Teddy Bear's Picnic (TBP) is a nine item narrative story stem measure, designed to assess young children's internal perceptions of self and others (Mueller, 1996). The procedure was originally developed in America (Mueller & Tingley, 1990), then expanded and refined in the UK by Mueller. It is a semi-structured task, using a family of toy teddy bears, and a variety of props to illustrate and enact a picnic outing. Nine incomplete story stems are presented but left unresolved, and children are asked to complete each story in response to a prompt. The TBP attempts to reveal personal, emotionally related themes (constructs) in children aged four to six years.

1.5.2.7 Administration of the TBP

The TBP is administered to children individually using specific toys and props, and is video or audio recorded for later coding. Each story stem is presented using a standardised description of the scenario, followed by one of two prompts " what

happens next?” or “what does your bear do or say?”. At the outset, children are introduced to their ‘special bear’ through completion of a practice story using the same sex bear. Following completion of the practice stem, and prior to delivery of the first full story, the three remaining bears from the family are shown to the child, who is asked to identify each character. At this stage, the examiner establishes how the child refers to his/her special bear. Administration of the TBP takes between twenty and thirty minutes.

1.5.2.8 Themes identified in the TBP

A total of nineteen themes are identified throughout the TBP. Valuing relates to the main character valuing others, and includes any affiliative expression or plan directed towards another character in the story. This theme also includes main character compliance with appropriate/legitimate parental sanction. It is a theme relating to prosocial acts and behaviours initiated by the main character. Devaluing relates to the main character devaluing others. This includes any act directed towards another which is unkind, mean, or suggests they are worthless. It also includes defiant behaviour which is not aggressive. Corresponding themes exist relating to the main character experiencing valuing or devaluing actions or behaviours from others towards him or herself. These include praising, helping, nurturing, supportive and caring acts directed towards the main character; and rejection, disproportionate verbal punishment, sense of worthlessness and/or badness respectively. Devaluing themes are associated with figures of authority i.e. parents, and teacher (or other adults) rather than siblings and peers. Drive expression themes include aggression, sexual and/or oral expression, and physical punishment that is out of proportion to the event. This theme includes all characters in the story, and most commonly relates to aggression but also includes other drive expression and prolonged self indulgence by any character. Main character resourcefulness is a theme relating to motivation, determination and confidence, and is directly related to the main character’s ability to solve practical and emotional problems. Resourcefulness in others fulfils the same criteria but is attributed to other characters in the story. Themes relating to vulnerability in the main character and vulnerability in others (or physical objects) include an inability to successfully manage situations which arise, and includes fear, helplessness and physical symptoms of sickness or excessive tiredness. Themes relating to emotional expression include all spontaneous expressions of positive or negative affect in the main character and others. The emotional expression theme also takes note of the

affective quality of the story ending. Additional themes relate to the whole story or any part of it, and include disorientation or confusion in the story, immaturity, and distraction/impulsivity at any stage across the course of the test.

1.5.2.9 Coding the TBP

The TBP is coded from observable actions, comments or words used during the interactions between the bears during the stories generated by the child. All 19 item themes are scored according to their presence or absence in each story stem. 2 themes are coded but not included in the total negative score. These are immature story and distraction/impulsivity, both of which are described as relevant to the reliability of the final total scores, but do not contribute in their own right.

Studies which have used narrative measures

Several studies have now begun to illustrate how useful narrative story stem techniques are in illuminating psychological processes in young children. The following seven studies have been administered with young children in a variety of settings, and include examples of how the technique can be used with children who are identified as having difficulties, as well as with non clinical samples.

1.5.3 Studies using the MSSB

Five studies have been selected, published since 1990, which have used the MSSB as a direct measure of children's narrative expression.

1.5.3.1 Play narratives in 36 month old children. Early moral development and family relationships. Buchsbaum & Emde, (1990).

This study set out to elicit representations of affective themes relating to moral conflicts in 36 month old children. The idea for the study stemmed from debate between psychoanalytic theorists about the presence or absence of significant moral development in children prior to the resolution of the oedipal conflict (at about 4 years). In order to elicit relevant representations to test out the hypothesis that moral development does occur prior to 4 years, the authors devised a story stem technique, based on play narratives, to explore 3 year old children's responses to moral dilemmas. It is based on a model which supports the view that early moral development has its origin in biological preparedness and experiences of early caregiving relationships. This study reports the very early development and application of the MSSB.

26 first born children (16 girls, 10 boys) took part in the study. All parents were middle class, white, and married. More than half were from dual career families. Four narrative story stems were completed with each child in the home setting, and eleven in the laboratory. All material was videotaped for later coding.

Two areas were explored using stories devised to focus on specific themes relating to moral conflict and family relationships. Moral development was represented through stems which focused on empathic or prosocial responses and rule adherence. Family relationships were represented through stems which focused on reparation and separation responses. The stories were coded according to thematic content, and style of conflict resolution. All children completed all the stems in the home setting, and all but one in the laboratory setting, who was described by the investigators as 'competent but oppositional'. Each story was found to prompt a specific range of responses. The authors suggest that the limited number of coherent stories in response to each stem demonstrates that the children responded by organising their knowledge and experience in direct response to specified affective themes. For some children, each story was not discretely resolved in that characters and themes would spill over into the next stem. This was hypothesised to be a reflection of the limited experience of others outside immediate care givers for children of this age. Findings demonstrated stability over time and adaptive variation across characters in how they resolved conflict. There was less stability in story endings which was suggested to be a direct response to very recent experience.

This early study provides preliminary evidence for moral development in children as young as 3 years. Children presented meaningful responses to story stems which sought prosocial/empathic outcomes, rule adherence and resolution of a moral dilemma. Examples of responses included sharing, helping, comforting, soothing (prosocial), doing as parent figures asked (rule adherence), and grappling with a prohibited prosocial behaviour (moral dilemma). In response to stems devised to explore family relationships, more children represented difficulty in separating from parent figures in the 'exclusion' story than 'departure and reunion', suggesting that accessible but emotionally unavailable caregiving poses significant difficulties for emotional management (Sorce & Emde, 1981).

The study focused entirely on a sample of middle class children whose experience

of caregiving relationships was largely similar. The authors acknowledge the need for further research with larger samples of children, including at risk and clinical groups.

1.5.3.1.1 Key points

The study provides empirical evidence to support the hypothesis that early moral development and emotional coherence can be tapped using a narrative based measure with very young children. The stories were specifically selected for use in the study from a range of available stems in the MSSB. This offers considerable flexibility in applying the test in both clinical and research settings - each story is a discrete assessment in its own right, and stories can therefore be selected (and developed, if necessary) to investigate specific themes and relationship representations in children as young as 3 years.

1.5.3.2 The use of a narrative story stem technique with maltreated children: Implications for theory and practice. Buchsbaum, Toth, Clyman, Cicchetti & Emde, (1992).

This study focused on increasing understanding of psychopathology in maltreated children. Exploration of the domains of emotion regulation, internal representations of relationships and moral development was facilitated using an adapted version of the MSSB in a sample of 100 four to five year old maltreated children, and a matched non-maltreated comparison group. Themes for exploration were based on previous research demonstrating an association between maltreatment in childhood and affect regulation (Cicchetti, Ganiban & Barnett, 1991), insecure attachment relationships (Carlson, Cicchetti, Barnett & Braunwald, 1989; Crittenden, 1988; Cicchetti & Barnett, 1991), aggressive behaviours (Kaufman & Cicchetti, 1989); and early moral development and emotion regulation (Emde et al., 1991).

100 children were selected from a clinical population to take part in the study. The children were aged between 4 and 5 years, and were already attending a family centre for intervention. Each child had a history of maltreatment, and had experienced either physical abuse, sexual abuse, emotional abuse, neglect or a combination of types of abuse. All were from low socioeconomic status backgrounds. A comparison group of non-maltreated children was recruited from lists of 'aid to families with dependent children', which acted as a marker for low socioeconomic status. Absence of a history of maltreatment was verified using

records from Social Services departments. The groups were individually matched for age, sex, race and receptive language (using the Peabody Picture Vocabulary Test - Revised; Dunn & Dunn, 1981). All children completed the same story telling task, and videotapes were used for later coding. The authors do not state if the interviewers were blind to which children were from the clinical sample.

Specific story stems were selected to explore representations of moral development (Buchsbaum & Emde 1990), and family relationships (Bretherton et al., 1990). A total of nine story stems were used.

A case study approach was used to report the findings, based on two representative cases of maltreated and non maltreated children from the sample (n=4). Responses from maltreated children were described as containing more themes involving inappropriate aggression, neglect and sexualized behaviours, generalised statements about self being bad, a lack of help being offered to an injured character, and aggressive, punitive and abusive language.

This study provides good insight from a clinical perspective into the usefulness of the MSSB narrative technique with young children. The content themes and representations of the maltreated and non-maltreated children are compared and contrasted in detail, and implications for assessment and intervention are discussed. The study lacks empirical data from which comparisons could have been made between children's adaptive and maladaptive responses, and various factors associated with their particular experience of maltreatment e.g. type of abuse, duration. The association between narrative response and language development is not tested at all, and significant risk and protective factors in each child's family background were not explored. Both issues raise questions about the validity of the assumption that all differences between the groups can be attributed to the effects of maltreatment. The process of administering the story stems is described in general terms only, and testers used 'non directive' prompts, as well as standard probes to encourage the child to resolve each hypothetical dilemma. This is likely to have been influenced by the testers background knowledge of the child's situation, and increased the possibility of over focusing/leading on some themes. However, several themes elicited from the maltreated children were consistent with processes and dynamics identified in previous research e.g. victim/victimiser role reversal (Troy & Sroufe, 1987), disruptive attachment relationships (Main & Hesse, 1990), and

disrupted patterns of emotion regulation (Aber & Cicchetti, 1984; Gaensbauer & Harmon, 1982).

1.5.3.2.1 Key points

As in the previous study, this research demonstrates the applicability of an adapted version of the MSSB with children of preschool age, where specific stories were selected to investigate specific theoretically defined themes. Material elicited from children with a history of maltreatment illustrates the value of engaging in story telling as a method for clinical assessment and intervention.

1.5.3.3 Can emotions and themes in children's play predict behaviour problems? Warren, Oppenheim & Emde (1996).

This study empirically tested the relationship between emotions and themes in children's play, and teacher and parent rated behaviour problems in a group of fifty one children aged three to seven years. The study had two main purposes. Firstly, the authors wanted to explore the usefulness of the MSSB in providing a standardised self report assessment tool for children, and secondly to determine whether children's individual representations of thematic material could be useful in clinical assessment. Markers of disturbance (distress and aggressive/destructive themes) provided the focus of the study, compared with standardised ratings of behaviour problems.

51 children were recruited using a voluntary opt-in procedure in a non clinical population. All mothers were already involved in a larger family study about children's emotional development.

Three groups of measures were administered at three time points (children aged 41 months, 55 months and 66 months). Behaviour problems were measured using the parent completed Child Behaviour Checklist (CBCL), and teacher completed Teacher Report Form (TRF) (Achenbach, 1991). The Peabody Picture Vocabulary Test - Revised (Dunn & Dunn, 1981), the Expressive One - Word Picture Vocabulary Test - Revised (Gardner, 1990), and the vocabulary sub test of the Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1967) were used to assess language development.

Twelve story stems from the MSSB were administered to each child on an individual

basis at each time point. Videos were coded at a later stage by narrative coders who were blind to the hypotheses.

Results demonstrated a complex association between themes relating to distress and destructiveness in four to five year old play, and parent and teacher rated behaviour problems. These themes did not show stability between three and four years, in contrast to internalising and externalising ratings on the CBCL.

Themes relating to distress correlated significantly with CBCL and TRF when expressed during the narrative phase of the MSSB, but not during the presentation and transition phases, where mean level of distress was low. Themes relating to anger, sadness and concern did not significantly correlate with CBCL or TRF. Aggressive/destructive themes correlated significantly with CBCL and TRF at age four to five years. Partial correlations were calculated which demonstrated that individual differences in temperament, language ability and investment in the task did not account for the differences in children's themes. Narrative coherence increased with age, where more than half of the children at age three years could not complete a coherent story for more than half of the time. This was not the case for any of the older children. Similarly, four year old children were more involved in telling stories than three year old, as measured by the 'investment in performance' code.

Findings from the study provide empirical evidence to support the hypothesis that themes and emotions in children's play are associated with childhood problems. The study demonstrates a link between distress and destructive themes during the narrative phase of the interaction and behaviour problems as measured by parents and teachers in four and five year old children. These findings are important in identifying specific age differences between children in their ability to create narrative stories, perhaps suggesting that both language development, and themes relating to emotional dysregulation are in transition between three and four years, becoming more stable between the ages of four and five (Cole, Michel & O'Donnell-Teti, 1994). The association between externalising problems and themes identified during the stories provides promising information about the usefulness of the technique as a self report measure in young children. The study sample consisted of a low risk non-clinical group who volunteered to take part. Further research is needed with clinical samples, and in increasing understanding of the association

between narrative themes and internalising disorders.

1.5.3.3.1 Key points

This study demonstrates empirical links between coded themes from the MSSB and parent and teacher rated behaviour problems. The internal consistency of the aggregated 'destructive' theme is reported ($\alpha = .87$ at 4 years, $.84$ at 5 years); and inter-rater reliability based on intraclass correlations for emotional themes is also high ($r = .96$ for distress, $.81$ for anger, $.85$ for sadness, $.91$ for concern, and $.80$ for investment in performance).

1.5.3.4 Children's narratives about hypothetical moral dilemmas and objective measures of their conscience: Mutual relations and socialisation antecedents. Kochanska, Padavich & Koenig (1996).

The study examined the correspondence between measures of children's conscience, and observed and mother report of moral conduct. The authors aimed to provide empirical data to support the assumption that children's responses to stories in a modified version of the MSSB reflect their moral development (Buchsbaum & Emde, 1990; Dunn, Brown & Maguire, 1995). Two main goals were defined, to determine the empirical relationship between children's 'declarative knowledge' (as expressed in hypothetical narratives) and 'procedural knowledge' (observed actions) (Emde et al., 1991); and secondly, to explore the quality of parental discipline and its association with child conscience.

Data were collected as part of a larger longitudinal study of conscience development. 103 mothers and children were assessed at two time points in home and laboratory settings (26-41 months, 43-58 months). Videotaped observation in naturalistic and structured contexts, questionnaire data from mothers and children's narrative measures of conscience formed the data set.

Observations of moral conduct took the form of mother - child interactive discipline, which focused on typical issues from early childhood e.g. clearing away toys, not touching prohibited objects on a shelf, mealtime behaviour, play situations (Kochanska & Aksan, 1995). A total of 80 minutes of mother - child interactions at home, and 50 minutes in the laboratory were coded at time 1, and 55 minutes in the laboratory at time 2. Observations of maternal power assertions were coded from the same discipline contexts as those used for moral conduct. A detailed

observational coding system was used to collect data within each session.

Mothers completed two questionnaires at both time points, The Conscience Instrument (Kochanska et al., 1994), a hundred item questionnaire with multiple conscience scales; and Rothbart's Aggression Scale (Rothbart, Ahadi & Hershey, 1994). The CBCL (Achenbach, 1991) was completed at time two. Children's narrative measures of conscience were collected at time two. Six story stems were administered in a similar format to the MSSB. The story stems were modified and extended versions of story stems from the MSSB. The coding system was also redefined.

Children who described several antisocial narrative themes demonstrated more behaviour problems when compared with children who described several themes relating to commitment to/concern about good behaviour. Children who experienced more power - assertive discipline from their mothers described fewer narrative themes relating to commitment to/concern about good behaviour. These children were also rated higher on externalising problems using observed and mother reported measures at both time points. Girls demonstrated higher rates of emotional themes on narrative and objective measures of conscience.

This study provides empirical support for an association between children's narrative interpretations of hypothetical moral dilemmas and objective measures of their conscience. Children's declarative knowledge, as expressed in their narratives, was associated with past and current procedural action tendencies and with maternal reports. Maternal discipline style was associated with children's conscience development, specifically that maternal power - assertive discipline is detrimental to children's conscience development.

Modification of the MSSB related mainly to the administration of the battery. Additional instructions, questions and prompts were presented to the children throughout the procedure in a standardised form. Each child was invited to "tell stories about little people", and to select and name the protagonist "to be in all the stories". Following presentation of the stem, a standard series of questions and prompts (not reported) were asked about the protagonist's course of action and resolution of the situation, and the child's own feelings and actions taken if they were the protagonist in the presented situation. The responses of the examiner to

questions asked by the child throughout the presentation of the stems were also standardised. This is significantly more specific and direct than the procedure followed in the original form of the MSSB. The authors also modified the original coding system. All data were coded from videotape, and coders marked each theme that occurred during every thirty second period of each narrative. This was reported to facilitate the coding process and increase reliability.

This study provides important evidence in the area of early conscience development, and in the usefulness of narrative measures.

1.5.3.4.1 Key points

The administration and scoring of the MSSB was refined for use in this study. Standardised questions were used to elicit content and emotion themes, and standardised replies were defined for use in answering children's questions. These are likely to have increased consistency of delivery. The scoring of timed sections of video material is likely to have reduced the margin of error in coding, and increased scoring consistency.

1.5.3.5 Emotion regulation in mother-child narrative co-construction: Associations with children's narratives and adaptation. Oppenheim, Nir, Warren & Emde (1997).

This study used the MSSB to examine the association between early mother-child co-construction of a separation/reunion narrative, and children's concurrent and later emotion narratives and behaviour problems. The interpersonal context of narrative skill development, and the links already found between early mother-child conversation and narrative development formed the focus for the study (Reese & Fivish, 1993; Fivush, 1991; McCabe & Peterson, 1991; Hudson, 1990). The first aim was to explore the association between early mother-child co-constructions and concurrent and later narratives about relationships and emotion. Secondly, variations in early mother-child co-constructions and concurrent and later emotion and behaviour regulation was explored using a standardised behaviour checklist at two time points.

51 low risk mothers and children volunteered as part of a larger longitudinal study of preschool children's narrative development in the context of the family. Children were assessed at two time points, 4 years 7 months and one year later. Children

and mothers carried out an observed separation/reunion co-construction in a laboratory setting. Free play was followed with guidance for mothers and children in using the materials to tell the story of a mum and dad who go on a trip without their son or daughter, then return later. Videotapes were used for later coding. Tapes were coded by two independent coders. Three segments of the interaction were coded, the separation phase, the parental absence phase and the reunion. Children's emotional coherence and maternal facilitation were rated using scales developed by the authors. Each consisted of 5 points with behavioural anchors to increase reliability. Themes were similar to those found in the Adult Attachment Interview (Collins & Read, 1990). 11 story stems from the MSSB battery were administered and coded. The Expressive One Word Picture Vocabulary Test (EOWPVT) (Gardner, 1990) and the CBCL (Achenbach, 1991) were also administered.

Results suggest that child MSSB variables, EOWPVT and the internalising and externalising behaviour problems in the CBCL show significant continuity across time. Children who were more emotionally coherent during the mother/child co-constructions also had MSSB narratives that were more coherent, more prosocial themes and fewer aggressive themes at both time points. These children had fewer externalising behaviour problems at both ages. Children's emotional coherence during the co-constructions, and prosocial themes during the MSSB were not related to concurrent level of expressive vocabulary. Maternal facilitation of the interaction was not correlated with narrative themes in the MSSB. The study provides important evidence about the continuity and stability of themes derived from the MSSB and their association with behaviour problems.

Several findings are important. Firstly, children who were rated as more emotionally coherent during the co-constructions with their mothers were also able to engage in more coherent narratives when engaged in the MSSB with an interviewer. This suggests that their emotional coherence in the dyadic co-construction is a necessary prerequisite for independent narrative coherence. Secondly, the results showed important links between children's emotionally coherent narrative ability and behavioural and emotional regulation as rated by their mothers. The associations were primarily found between themes in the MSSB and externalising behaviour problems. This could be due to the sample used in the study (white middle class), who were rated as experiencing very few internalising problems. However, this

requires further investigation as one strength of a narrative method is that it is able to tap children's internal experience, as well as external manifestations of problems, and this was not evident in this study.

1.5.3.5.1 Key points

This study provides evidence to suggest that children's language development is not associated with positive themes generated from the MSSB, which in turn are empirically linked to emotional coherence and the absence of 'externalising' problems. The study demonstrates the importance of mother-child co-construction in the development of emotional coherence.

Narrative coding was based on the system developed by the MacArthur Narrative Working Group (Warren et al., 1993). Content themes were coded according to categories (empathy/helping; reparation/guilt; affiliation; affection; interpersonal conflict resolution; aggression; atypical negative responses; personal injury). Each narrative event received one content theme, but because narratives contained several events, each whole narrative could be coded with several event based content themes. Total number of content themes were counted across all narratives and mean content theme scores were generated. Narrative coherence was coded on a 10 point continuous scale for each story stem, and added together across all stories to form a total score.

Internal consistency was reported as .87 for the narrative composite score, and inter-rater reliability was also high. To address the specific research question, two content theme composite scores were calculated - a prosocial themes composite score, and an aggressive themes composite score. Internal consistency is not reported for the content theme scores.

1.5.3.6 Summary of findings from US studies

US studies provide substantial evidence for the reliability and validity of the MSSB narrative measure. It's strengths lie in it's adaptability, where specific stories can be selected to address specific questions. The scoring system appears to generate reliable and valid composite scores in the studies where this is described, but the process of coding is detailed, complex, and highly labour intensive. One study (Kochanska et al., 1996) promotes the use of standardised questions to elicit responses from children, which are scored from videotapes in the form of frequency

counts within specified time intervals. This appears promising as a method which has the potential to increase reliability and validity.

1.5.4 Studies using the TBP

Two studies have been published which illustrate the development of the TBP.

1.5.4.1 The Bear's Picnic: Children's representations of themselves and their families. Mueller & Tingley (1990).

This study was designed to elicit fantasy story telling material from young children to increase understanding of their core representations of self and their family, by accessing material which is outside conscious awareness and therefore not directly available for assessment. The study provides early empirical data using the Bear's Picnic research instrument, which uses doll play and story stems to reveal representations of self and other. The method is derived from a combined theoretical base, described by Mueller (1996) as a 'contemporary cognitive approach'. The approach essentially views children's social development as a form of social 'scheme' (Bee, 1992), where schemes are defined as relatively enduring internal theories, attitudes and beliefs through which sense is made of the actions of others. Parallels can be drawn with the 'internal working models of self and others' in attachment theory (Bowlby, 1988), the distinctions between procedural and declarative knowledge and the development of consciousness in the psychoanalytic literature (Emde et al., 1991); and personal construct theory (Kelly, 1955) which emphasises the importance of an individual's internally defined 'personal constructs' in developing an understanding of the world. The Bear's Picnic is therefore a narrative measure which seeks to tap internal representations of self and other through the use of a story-telling, play based technique.

27 children (14 girls, 13 boys) from middle class backgrounds participated in the study at age 4 years, all of whom were part of a larger longitudinal study of temperament. Each child in the sample had been assessed at 20 months interacting in a play session with their mother. This videotaped material was used in the Bear's Picnic study.

Six story stems were presented to each child, following an introduction of the technique and completion of a warm up story. All stories involved representations of the self in relation to others in the family. The stories were all completed in the home

setting by a female examiner who was blind to the findings from the mother - child play session 28 months earlier. The Bear's Picnic was administered individually with no other family members present, and videotaped for later coding.

Firstly, tapes from the dyadic play session were analysed for quality of mother - child interaction. A seven point rating scale was devised for use across four dimensions. These consisted of maternal sensitivity, sharing of control, positive or negative child affect, and maternal affect. A composite score was derived for each dyad on each dimension, and inter-rater agreement was high (.77 - .79). The Bear's Picnic stories were then analysed for content, and common themes and typical responses were recorded. Three 'markers' were initially scored. These were fantasy failure (lack of ability to produce a fantasy), positive valuing of self and other, and negative valuing of self and other, based on the central hypothesis that positive and negative statements about characters within the stories reveal the child's inner model of self and other. Inter-rater agreement was high (.79 for total positive scores, .89 for total negative scores).

Results revealed two significant associations between mother-child interaction at 20 months and Bear's Picnic representations at 4 years. Maternal sensitivity at 20 months was negatively associated with the frequency of devaluing themes relating to self/other in the Bear's Picnic. Maternal positive affect at 20 months was negatively associated with the frequency of devaluing themes in the Bear's Picnic for boys only. Five additional associations approached significance.

Three dimensions of family relationships were revealed through the Bear's Picnic. These were summarised as themes relating to engaged/unengaged in family life, equal/dominant relationships within the family, and cooperative/conflictual relationships.

This study provides important preliminary evidence to support the validity of the technique in eliciting children's representations of self and their family. The composite scoring method, focusing on the frequency of valuing (positive) and devaluing (negative) representations resulted in high inter-rater reliability, which lends support to the theoretical model underpinning the development of the Bear's Picnic.

The study demonstrates that maternal sensitivity at 20 months was associated with children's positive/negative representations of self and other at 4 years. Differences between boys and girls in respect of findings relating to maternal affect require further investigation. Additional associations which approached significance are important as the sample was very small and not representative of the population.

1.5.4.1.1 Key points

Three themes in total were coded in this early version of the measure, fantasy failure, positive valuing of self and others, and negative valuing of self and others (later defined as devaluing). Positive and negative valuing of self and other were each combined into single themes and each theme was scored as present for its first occurrence only i.e. the maximum score per theme per story was 1. The total number of positive and negative codes were added together, and inter-rater reliability coefficients of .79 and .89 were derived. A new code, 'quality of family relationships' was developed from children's responses in the study, but not reported from this sample.

Correlation coefficients were calculated to test the association between the 3 themes gathered from children at 4 years and mother-child interaction at 20 months. 12 correlations were calculated. 2 reached significance, 5 were in the expected direction but not significant, and 5 were not reported at all. Small sample size is described as the main obstacle ($n = 27$). Internal consistency of total scores was not reported.

1.5.4.2 The Teddy Bear's Picnic: Four year old children's personal constructs in relation to behavioural problems and to teacher global concern. Mueller (1996).

This study was designed to establish concurrent validity for the TBP by comparing findings from coded stories with a validated teacher completed behaviour checklist. The TBP is a refined and expanded version of the Bear's Picnic, during which children are asked to complete nine stories about a family of toy bears who go out for a picnic. There were two parts to the study. Firstly, the study aimed to demonstrate that TBP could discriminate between children above and below cut-off for the likely presence of emotional and behavioural problems on the standardised teacher measure. Secondly, teachers were asked to rate their level of global concern about each child's behaviour or adjustment. The study examined the

separate and combined ability of the TBP and teacher behaviour measures to discriminate between these groups.

51 children attending nursery school in an area of high social disadvantage took part in the study. Children were aged between 4 years and 4 years 8 months. Parents were asked for permission to include their child in the study, and all parents agreed.

Children were assessed individually in a separate room at the nursery by one of three trained testers. All stories were tape recorded for later coding. Testing was completed within a three week period, and during this period, the nursery teacher best acquainted with the child completed the Preschool Behaviour Checklist (PBCL) (McGuire & Richman, 1986). At the end of the PBCL, teachers were asked to rate their level of concern about the child in answer to the question "Are you concerned about this child's behaviour or adjustment?" Teacher's answered by ticking one of three boxes, labelled 'no concern', 'possible concern', and 'yes, definite concern'. (McGuire & Richman, 1986). The TBP was administered without knowledge of the PBCL or level of teacher rated concern.

Findings demonstrated that the TBP could discriminate between children who were above and below cut-off on the teacher measure. In addition, both measures were able to distinguish between levels of teacher rated concern about behaviour or adjustment. Combined results from both tests were more powerful in distinguishing level of teacher concern than either test on it's own. The study concluded that the TBP is a useful measure of children's perceptions of self and others, and that these perceptions are systematically related to children's mental health.

The study provides empirical evidence which supports the usefulness of the TBP narrative measure with young children. Codes were based on tape recordings of the procedure, suggesting that non verbal actions and gestures were not included in the coding system. Reference is made to the importance of 'acting out' stories using little or no connected speech, and how this information was reliably used in the coding process is unclear. In addition, the study design did not include measures of concurrent language ability, thus preventing analysis of the association between verbal ability and performance on the TBP. However previous studies have demonstrated that narrative themes and expressive language ability are not associated (e.g. Warren et al., 1996). The results provide some evidence to suggest

that teachers were able to recognise internalising as well as externalising problems in children, in contrast to earlier findings which support the hypothesis that teachers are more accurate in predicting non compliance and aggression, and parents are better at pinpointing isolation and withdrawal (e.g. Hinshaw, Han, Erhardt & Huber, 1992).

TBP findings from a small sample of children ($n = 4$) are described in detail to illustrate examples of misclassification. Each of these children's narratives led to high scores on individual TBP items which were strongly suggestive of clinical problems, yet were often masked by the use of summary composite scores. Several areas are suggested for further research.

1.5.4.2.1 Key points

This study generates factor based subscale scores based on factor analysis of 10 of the possible 17 scored codes. Several points are relevant. Two new stories were used in the TBP in addition to the 7 from the Bear's Picnic. These were the separation/reunion story delivered at the end of the TBP, and the classroom based 'chalkboard' story. An additional 14 item codes were developed in addition to the 3 from the Bear's Picnic. The valuing/devaluing codes were transformed into separate rather than composite codes, and the 'fantasy failure' code is renamed 'immature story', and extended in definition to include a range of immature responses from children. 17 codes (6 positive codes and 11 negative codes) were summed to generate positive and negative total scores. The 'immature story' and newly developed 'child distracted/impulsive' codes were not included in the negative total score.

The study does not discuss the development of the measure from its early Bear's Picnic version to the TBP version. However, the new version is empirically supported on the basis of strong inter-rater reliability, and concurrent and discriminant validity results. The internal consistency of the factor based subscale scores is not reported.

The main purpose of the study related to establishing validity of the measure within the school context, using concurrent teacher behaviour ratings. Mueller (1996) focuses on the need to extend research using the same concurrent measures with larger samples of 'troubled' and 'normal' children to:

- establish which children are misclassified by each test, and
- establish the meaning of high single item scores.

These, and additional questions will be considered in the current study.

1.5.4.3 Summary of UK studies

The development of narrative measures in the UK has focused predominantly on this research group, who have developed the early Bear's Picnic and subsequent Teddy Bear's Picnic measures. The measure uses stories which tap a broad range of psychological themes relevant to the internal perceptions of the child. The coding of material generated from the stories is based on single occurrences of themes which are represented in children's stories, and total scores are used to summarise the expression of positive and negative themes. The TBP has been validated through the establishment of concurrent validity with a teacher behaviour measure and rating of global concern, and appears promising as a new direct measure for use with young children.

1.5.5 Summary

Child mental health problems have, historically, been identified and reported by adult informants from the two main settings in which children develop - home and school. The prevention of problems in early childhood has become an increasing focus in clinical psychology as epidemiological studies have provided strong evidence to demonstrate links between the onset of difficulties in early childhood and the continuity and comorbidity of disorders in later childhood, adolescence and adulthood. The development of the capacity for emotion regulation is important in understanding the onset and maintenance of emotional and behavioural problems in childhood, and is predictive of adolescent and adult psychological adjustment. Reliable and valid measures of mental health status in young children have tended to rely on observation and adult informant based interviews and questionnaires. Measures used directly with children have focused almost entirely on motor development in the pre and early school years, and as a consequence, children themselves have rarely contributed to their own assessment in clinical settings. However, there is considerable evidence to suggest that reliance on observable actions and behaviours pays too little attention to the child's internal representations of self and others.

Measures which focus on children's narrative expression have recently been developed, and a growing number of studies support the ability of the method to elicit core internal information from children in a non threatening, age appropriate manner. Two similar techniques are currently being studied. The MSSB, devised in the US, has been evaluated in a number of studies designed to test hypotheses relating to children's development and early care experiences. The TBP, developed predominantly in the UK, has focused on the association between themes generated in children's stories, and teacher reported behaviour problems.

1.5.6 Overview of the study

The current study aims to extend the work of Mueller (1996) in three areas. First, reliability analyses will be carried out on scores generated from the TBP to include measures of inter-rater reliability and internal consistency. Second, information will be collected concurrently from parents as well as teachers, using corresponding ratings of children's emotion and behaviour, and global concern. Parents are an important source of information about children's mental health, and as highlighted by Offord et al (1996), perceptions of children's strengths and difficulties are informant and setting specific. Third, the study investigates the ability of the TBP to correctly classify children into a 'borderline clinical' group, as well as 'clinical' and 'normal' groups, based on parent and teacher ratings of emotional and behavioural difficulties.

As investigated by Mueller (1996), this study examines the ability of the TBP to correctly classify children into groups based on ratings of global concern. In addition to teacher ratings of global concern, this study also considers parent rated concern. The TBP may be used in two ways, either:

- 1) as a screening tool to identify children who are at risk for, or already experiencing the early signs of psychological adjustment problems, or
- 2) as a clinical measure used to explore children's perspectives of self and others.

The study aims to examine the validity of the measure for both purposes. The usefulness of using the TBP with individual children will be explored.

1.5.7 Hypotheses

The study hypothesises that:

- 1. TBP total positive and total negative theme scores will discriminate between children who are rated as above and below clinical cut-off score on a parent and teacher behaviour measure (identifying 'cases')**

- 2. TBP total positive and total negative theme scores will discriminate between children who are rated as above and below the 'borderline' clinical cut-off score on a parent and teacher behaviour measure (identifying children who are at risk of developing emotional or behavioural problems)**

- 3. TBP total positive and total negative theme scores will discriminate between levels of parent concern about their children's behaviour**

- 4. TBP total positive and total negative theme scores will discriminate between levels of teacher concern about child behaviour.**

2. Method

2.1 Design

In order to test out the hypotheses, a cross sectional study was carried out with one group of children from a community population. This group was already participating in a larger population based controlled trial (Appleton, Whitaker, Wilkinson, Hibbs & Edwards, 1997; Hammond-Rowley & Appleton, 1997).

2.2 Sample

All children from one reception intake at one primary school were selected to take part in the study (2 classes, n=71; see Figure 1). This group was selected for four main reasons:

- the school was already included in a larger population based controlled trial within the same community (Appleton et al., 1997). Baseline parent and teacher measures of behaviour and ratings of global concern were therefore already being collected at a specific time point, thus preventing the need for additional data collection from adult informants
- all children from one year group were selected in preference to a random sample to avoid the potential that parents and teachers may feel that children had been specifically selected, perhaps suggesting that children with problems were the main research interest
- post baseline 'screen', this school was not going to receive a classroom based direct intervention in social and emotional problem solving (Greenberg & Kusch, 1995). Delivery of this intervention was planned in other schools within the same community. It was important not to raise concerns in parents and teachers of children who were to receive the classroom based intervention that findings from the study would be used to inform the intervention process; and
- the TBP measure was designed for use with young children aged 4-6 years 11 months. Reception age children are within this age range.

Additional practical reasons for selecting this school included distance from base, already established relationships with staff as a result of the larger study (Fisher,

1995), and appropriate size of year group within the parameters of the study.

2.2.1 Criteria for inclusion in TBP study

The following inclusion criteria were defined:

- child attending either reception class in the identified school
- child resident in the intervention community within the population based controlled trial
- child's parents had completed the parent questionnaires as part of the larger study
- child's teacher had completed the teacher questionnaires as part of the larger study
- child's parent(s) did not refuse permission for them to take part.

2.2.2 Total sample meeting criteria for inclusion

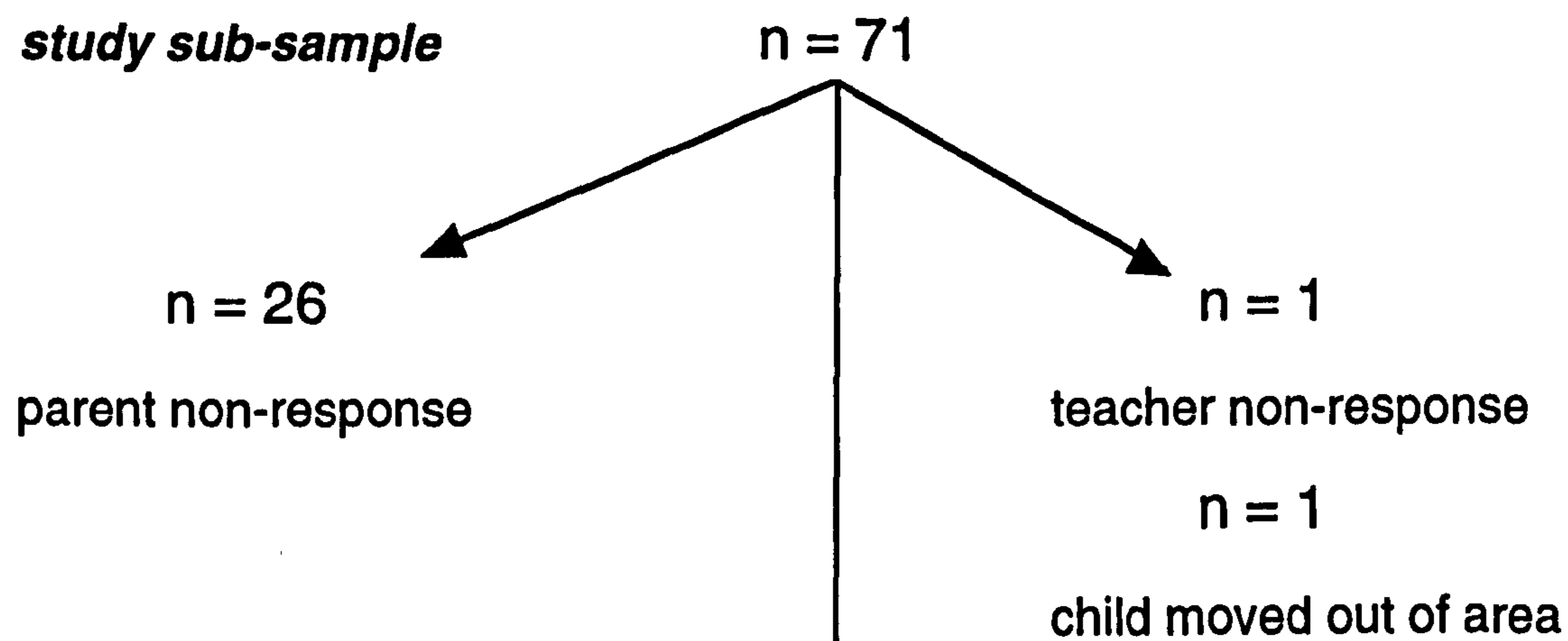
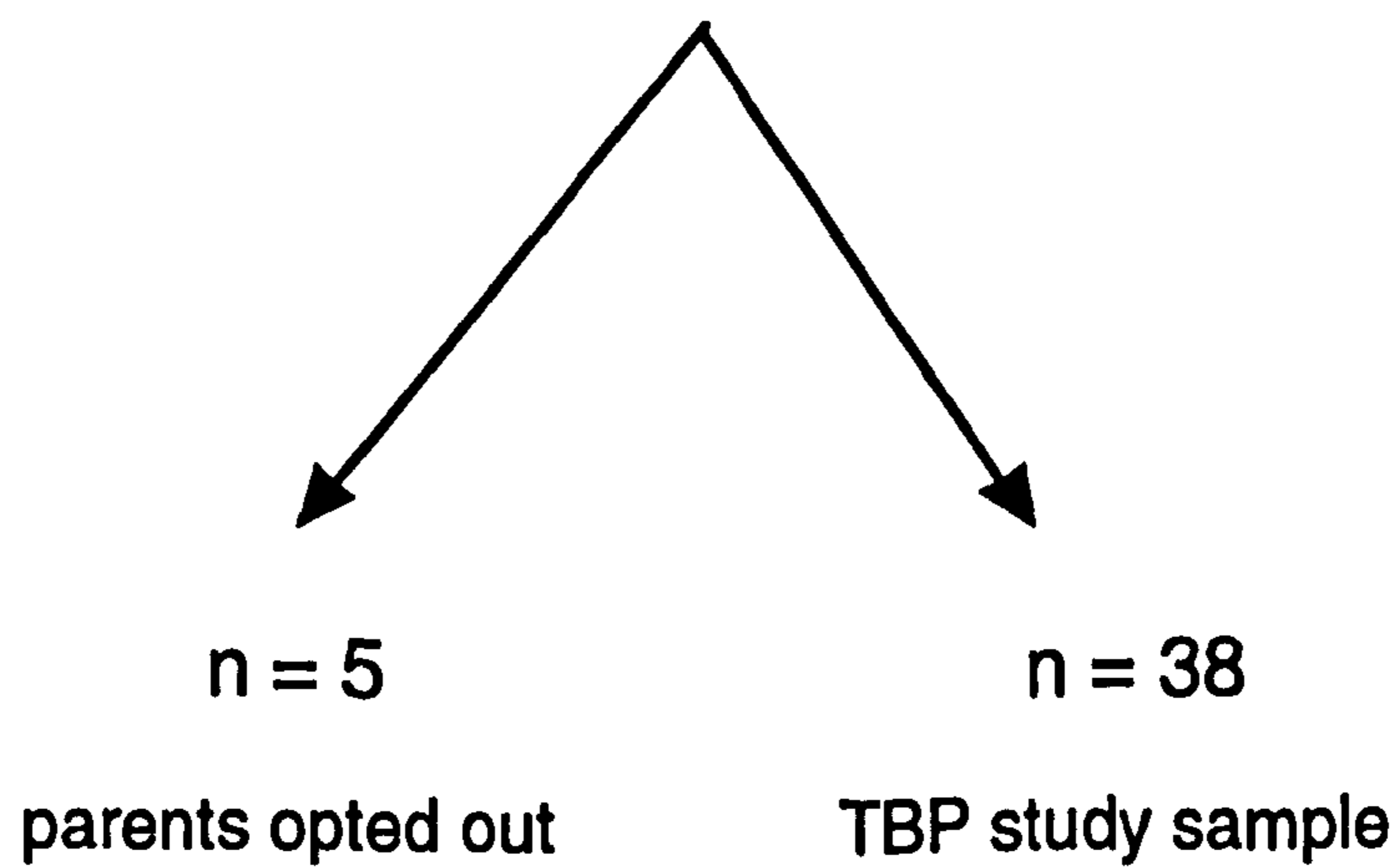
From the total sample of 71, one child had moved out of the area and hence no longer attended the school. 26 children's parents did not return the questionnaires within the baseline 'screen', and 1 child's teacher did not complete the teacher questionnaire. 43 children met the first stage criteria for inclusion. From this sample of 43, 5 parents refused permission for their child to take part, resulting in a total final sample of 38. This constituted 56% of the 2 classes. Figure 1 over leaf illustrates the 2 stage selection procedure.

2.3 Measures

2.3.1 TBP

The TBP is described in the introduction sections 1.5.2.6 - 1.5.2.9. Nine story stems form the complete version of the TBP. The test reveals concurrent psychosocial constructs through the telling and 'acting out' of stories, rather than the specific events which may have led to the formation of these constructs. The following examples illustrate two story stems in their full version. The full set of story stems is available in Appendix 1.

Figure 1.

TBP study sample***Controlled trial******study sub-sample******TBP potential
study sample***

Story 1

Props: whole bear family, picnic blanket, toy wagon, tree

Stem: “One day, the bear family went for a picnic. Mummy and daddy bear are over here setting up the picnic. (Demonstrate with toys and props). Your special bear (name) is over here playing with the wagon (demonstrate) when suddenly, he/she hears a noise and runs over here to see what it is (demonstrate bear running to tree). It’s just a little squirrel, so your special bear (name) goes back to playing with the wagon. But when he/she gets back, look what’s happened! Sister/brother bear has taken the wagon! (Demonstrate). What happens next?”

Themes sought: Perception of the quality of sibling relationships; co-operative play vs punitiveness and fighting.

Story 2

Props: one small rock

Stem: “After the picnic, your bear (name) and sister/brother bear go for a long walk in the woods, way way over here (Demonstrate, moving bears as far away as possible from the parent bears). Suddenly, sister/brother bear trips over a rock (demonstrate) and hurts her/his leg, and can’t get up. What does your bear (name) do?”

Themes sought: Coping strategies; resourcefulness.

Themes sought from each story stem are presented in Table 1 over leaf.

2.3.2 Reliability and validity of the TBP

In the original study (Mueller, 1996), reliability and validity were examined based on summary scores derived from 7 of the total available 17 codes. Inter-rater reliability was established (mean $r = .87$). Concurrent validity was supported in two ways. Firstly, correlations between summary scores generated from the measure and teacher ratings on the PBCL (McGuire & Richman, 1986) revealed inter-correlations of moderate size, in the expected direction. Secondly, single and stepwise discriminant function analyses were used to determine the separate and combined abilities of the TBP and teacher measure to predict teacher concern ratings. Mueller (1996) asked teachers to make a judgment of concern about each child, using a question included in the PBCL (McGuire & Richman, 1986). The question asked “are you concerned about this child’s behaviour or adjustment?” Teachers responded by ticking boxes labelled “no”, “possibly”, or “yes”. Based on summary variables from both measures, and using ‘teacher concern group’ as the grouping variable, TBP effectively discriminated between children in two of the three teacher concern groups

PAGE/PAGES
EXCLUDED
UNDER
INSTRUCTION
FROM
UNIVERSITY

(“yes” and “no” responses). PBCL (McGuire & Richman, 1986) was a better discriminator than TBP overall, but neither test was powerful in correctly predicting which children were in the “probably” concerned group.

2.3.1.2 Factor analysis of the TBP (Mueller, 1996)

The 7 items which were used to generate total positive and negative sum scores were selected on the basis of a factor analysis, which used 10 of the total 17 available item codes from the measure. It is not clear how Mueller (1996) selected the 10 items which were entered into the factor analysis. These consisted of:

- main character values others
- main character devalues others
- valuing from others
- devaluing from others
- main character resourceful
- main character vulnerability
- vulnerability in others
- aggression in the story
- sad story ending
- story disoriented.

Factor analysis of these item codes using varimax rotation revealed a three factor structure. The first factor loaded high on four negative item codes (main character devalues others, vulnerability in others, aggression in the story and story disoriented). Devaluing from others also loaded on this factor, but did not meet minimum loading requirement as defined by the author (.40). This item was therefore excluded. The second factor loaded high on all three positive item codes (main character values others, valuing from others, and main character resourceful), and also on two negative item codes, which loaded in a positive direction (vulnerability in others and sad story ending). Both exceeded minimum loading requirement, but the first loaded higher on factor 1. Sad story ending was excluded. The third factor loaded high on two negative item codes (main character vulnerability and vulnerability in others). Vulnerability in others loaded higher on factor 1, and main character vulnerability was left as a single item in factor 3 and was also excluded. 7 items were therefore retained on the basis of this analysis.

Clinically relevant labels were attached to each factor on the basis of the pattern of item loadings (factor 1 'troubled child', factor 2 'ego resiliency' and factor 3 'vulnerability'). Summary variables were then generated, based on items retained from factors 1 and 2, where TBP negative sum score was collated by adding together the four item codes which loaded into factor 1 ('troubled child'), and TBP positive sum score was collated by adding together the three positive item codes which loaded into factor 2 ('ego resiliency'):

- Factor 1** Factor based subscale score = main character devalues others + vulnerability in others + aggression in the story + story disoriented
- Factor 2** Factor based subscale score = main character values others + valuing others + main character resourceful

2.3.2 Behaviour Measure

The parent and teacher versions of the Strengths and Difficulties Questionnaire (SDQ: Goodman, 1994b; 1997) were used as part of the larger study (see Appendix 2). The two versions are identical, allowing comparisons to be made between parent and teacher ratings of the same child. This is important, as findings in child mental health epidemiological research strongly suggest a low concordance between parent and teacher ratings of child psychological disorder (Rutter, 1989; Offord et al., 1996).

The SDQ is a brief behavioural screening instrument designed for children age 4 - 16 years. The questionnaire is based directly on the Rutter parent and teacher questionnaires which have well established reliability and validity (Rutter, 1967; Rutter, Tizard & Whitmore, 1970; Schacher, Rutter & Smith, 1981; Elander & Rutter, 1995). Goodman (1997) expanded the original Rutter questionnaire to include additional items relating to children's strengths. The SDQ questionnaire has been found to be highly correlated with the Rutter scores ($r = .9$) and has almost identical reliability and concurrent validity to those of the Rutter questionnaires (Goodman, 1997).

The SDQ asks informants about 25 attributes, 10 of which are generally described as strengths, 14 of which are generally described as difficulties, and 1 of which is neutral. Factor analysis has suggested that among children from non clinical samples, the expanded questionnaire is measuring five distinct dimensions:

PAGE/PAGES
EXCLUDED
UNDER
INSTRUCTION
FROM
UNIVERSITY

2.3.3 Rating of Concern

Parents and teachers were also asked to rate overall how concerned they were about the child's behaviour (see Appendix 3). Studies of child behaviour screening and outcome have established the importance of measuring impact of behaviour on the informant (Stallard, 1995; Goodman, 1997). Stallard (1995) suggests that there is no universal agreement about what constitutes 'problem' or difficult behaviour since all parents will view the same behaviour differently. In addition, previous experiences and expectations will influence parental perceptions of children's behaviour (Goodman, 1997), and parent concern is arguably a primary factor in assessing emotional and behavioural problems in children, as it is likely to be more relevant to the family context. The identification of behaviour problems from a parent defined rating of concern also addresses many of the flaws inherent within current classification procedures, and is likely to significantly reduce the reporting of false negative and false positive cases (Rutter et al., 1970). It is therefore likely to increase the early identification of problems at a stage before they become severe, which is central to the concept of prevention (Clements, 1985).

In this study, a simple categorical question, based on the work of Stallard (1995), was asked as part of the larger study in order to measure the impact of each child's behaviour on the informants. Parents were asked "overall, are you concerned or troubled about your child's behaviour?", and teachers were asked "overall, are you concerned or troubled about this child's behaviour?" Ratings of 'no concern', 'some concern' and 'definite concern' were generated by ticking the appropriate response box. Preliminary data from 188 parents shows that parent and teacher expression of concern is associated with higher SDQ scores (Appleton et al., 1997).

2.4 Procedure

2.4.1 Population based controlled trial

Data were collected in March 1998 from parents ($n = 1448$) and teachers in two matched communities, one receiving intervention and one acting as control, to establish baseline child mental health problem prevalence, and parenting confidence. (Appleton et al., 1997; Hammond-Rowley & Appleton, 1997). Three measures were completed by parents (SDQ Parent Version: Goodman 1997; Rating of Concern; Parenting Confidence Subscale: Family Experiences Questionnaire, Floyd & Zmich, 1991); the first two of which were used for this study. Teachers were asked to complete two measures on all children in their class at the same time (SDQ

Teacher Version: Goodman, 1997; Rating of Concern). Both teacher measures were used in this study.

H-R was directly involved in the organisation, planning and data collection during the preliminary child mental health pilot study and needs assessment, and in planning and delivering the interventions as part of the new service.

2.4.2 TBP Study

H-R was responsible for the study from the outset. Detailed negotiations were carried out with the school in the research planning stage, including piloting the measure with children from outside the research community, and acquiring Ethics approval from two committees. H-R supervised, trained and managed three colleagues (one Child Therapist, and two Research Assistants) in administering the measure, and coding the videotaped material.

2.4.2.1 Discussions with the school

A database was compiled of all children attending reception classes in the identified school, and checked against school lists. The Headteacher in the school was approached to ask permission to carry out the study, and to discuss the theoretical and clinical value of evaluating a direct measure with children, including the possible benefits for the school which may arise from such a study. Following an enthusiastic response from the Headteacher, meetings were held with the reception class teachers to seek their permission, explain the purpose of the study, and to address practical issues in relation to administering the TBP on an individual basis within the school environment. The classteachers expressed concerns about the videotaping of material, and felt that parents would be likely to opt out as a consequence of having concerns about confidentiality and data protection. It was therefore agreed that wherever possible, the video camera would be placed behind the children, so that the child's face was not directly in line with the camera, but that their actions and verbal responses would still be captured on tape. The classteachers also asked that the researchers spend some time in the classrooms prior to data collection so that the children would recognise them when carrying out the TBP on an individual basis.

2.4.2.2 Ethics approval

Following agreement with the school to proceed, ethics approval was sought from the University of Wales, Bangor Psychology Department; and the NWAHA (South

Clwyd) Ethics Committees. Permission was granted to proceed with the study, following amendment of the parent information letter to include more details to explain that information from the TBP would be compared with data collected in the larger controlled trial. Both committees agreed that an 'opt out' system for parent consent was acceptable on ethical grounds (see Appendix 4).

2.4.2.3 Piloting and training

Two standard families of toy bears were purchased for the TBP procedure (Marmalades Sylvanian Bear Family: Epoch Ltd, available from TOMY UK Ltd), and additional props were identified for use from existing sources. Formal piloting of the TBP was carried out by H-R with children of the same (or similar) age outside the research area, on a voluntary basis. Two research assistants were trained in the procedure (DC and JM), and practice sessions were organised within the service base between February and April 1998.

2.4.2.4 School visits

Visits to the school were organised at the request of the classteachers to familiarise the researchers with the school environment and the children. This took the form of joining in lessons, helping in the classroom and spending time with the teachers during break times to increase rapport, and finalise practical arrangements for the research phase. During this time, space was identified and adapted to ensure that the TBP could be carried out with minimum disruption to usual activities, whilst at the same time fulfilling school regulations which specify that adults do not spend time alone with children in areas where they cannot be observed.

2.4.2.5 Contacting parents

The final format of the letter to parents and 'opt out' slip were agreed with the school, and sent to all reception age children's parents via pupil post from the school on 23rd March 1998 (see Appendix 5). Parents were asked to complete and return the slip to their child's classteacher by the end of March if they did not want their child to take part. Parents were invited to contact H-R if they had any queries or questions about the study. Classteachers offered to canvass parents following receipt of the letter to ensure that all those who did not reply were aware of the study.

2.4.2.6 Sequence of research activities

A timetable of sessions during which the TBP would be administered was drawn up with the classteachers at the end of March, for data to be collected during April and May, so that all TBP assessments would be carried out within 6 weeks following completion of the parent and teacher questionnaires. Time was initially allocated based on maximum possible inclusion (n = 71).

As part of the larger controlled trial, all baseline parent questionnaires were sent out on March 16th 1998, and teacher questionnaires were delivered to the school on the same day. Parents returned their questionnaires from March 23rd through to the end of April, and completed teacher questionnaires were collected from the school on March 27th. The school contacted H-R by telephone on individual receipt of all TBP refusal slips, which were collected from the school at the end of the reply period.

Children who fulfilled all inclusion criteria for participation in the study were listed and double checked with baseline questionnaire returns and parent refusal slips. Dates of completion of the parent and teacher questionnaires were recorded, and an order for completion of the TBP was drawn up in order to minimise delay between the 3 measures.

Teachers were informed of the final list of children who were to take part in the study. Children were allocated to the timetable, and at the start of each day during the research, registers were checked to confirm that the identified children were at school. Adjustments were made if a child was off school by selecting the first child from the next research day to fill the allocated time. Table 3 over leaf summarises the sequence of research activities.

2.4.2.7 Administering the TBP

One researcher carried out the procedure at any one time, due to the limited availability of space within the school. All TBP assessments were video recorded for later coding. Video equipment and props for the procedure were set up in advance, with the camera positioned behind the chair in which the child would sit to maximise confidentiality, as agreed in the school, and explained to parents in the consent letter.

Table 3.

Sequence of research activity

| Date | Description | Personnel |
|------------------------------|--|--------------------------------|
| Nov 97 - Jan 98 | Detailed discussions with Headteacher and classteachers in one local primary school to plan the study | SH-R |
| Feb 98 - Mar 98 | Ethics approval sought from University of Wales, Bangor Psychology Department Ethics Committee | SH-R |
| | Ethics approval sought from NWAHA (South Clwyd) Ethics Committee | SH-R |
| | Pilot study of TBP with voluntary sample | SH-R |
| | Training with Child Therapist and Research Assistant | SH-R, DC, JM |
| | School visits | SH-R, DC, JM |
| | <i>Community Child Mental Health 'Screen' of all parents of children in nursery, reception and year 1 classes</i> | <i>Controlled trial</i> |
| Mar 98 - April 98 | 'Opt out' letters of consent for TBP study sent to all parents of children in reception classes in selected school | SH-R |
| | <i>'Screening' questionnaires returned to schools</i> | <i>Controlled trial</i> |
| | TBP parent 'opt out' slips returned to school | SH-R |
| | Timetable for TBP data collection finalised with school | SH-R, DC, JM |
| April 98 - May 98 | TBP data collection in school | DC, JM |
| June 98 | Thank you letters to parents and school | SH-R |
| July 98- Sept 98 | Coding of TBP video material Preliminary data analysis | SH-R, LM |
| Oct 98 | Dissemination of early findings to school and plans for extending research into TBP measure | SH-R |

Children were collected from their class, and asked to join the researcher for a short time to complete some stories. Children were then taken to the allocated space, which was positioned in the corridor in an area between the coat stands. This area is frequently used for individual work e.g. reading, so all children were familiar with the space. Spontaneous conversation was encouraged to allow the children to feel comfortable, and to explore the equipment. The presence of the video camera was explained, and children were told that they could stop at any time during the stories if they wanted to. The procedure was then completed in the standardised form as described above.

Following completion of the TBP, children were returned to their class by the researcher. Before collecting the next child, each video was labelled with the child's code and date, and stored in a separate box from the test equipment.

At the end of the study, letters were sent to all parents whose children had taken part in the study, and to the classteachers and Headteacher at the school to thank them for their help in participating in the study (see Appendix 6).

2.4.2.8 Coding the TBP

All TBP videos were scored at the service base using the coding manual for the procedure (see Appendix 7). Themes coded from each story stem are summarised in Table 1 above. Three videos were uncodeable due to errors in operating the camera (sound recording had been omitted on two, and the background noise interfered excessively on 1). 35 completed videos were coded. Coding was completed without knowledge of scores on the parent and teacher SDQ and concern ratings, which were collated by a different Research Assistant from the larger study. A random sample of 10 videos (29%) was independently coded by H-R, who was blind to the scores derived from the first coding, to check inter-rater reliability. A database was set up, and all data were entered in preparation for statistical analyses.

2.5 Statistical Analyses

Parent and teacher data, and TBP item codes were entered into a database using SPSS version 6.1 for Macintosh.

2.5.1 Exploratory analyses

Frequency distributions were generated for each of the main study variables, and mean scores were compared using one way analysis of variance (ANOVA).

2.5.2 Correlation coefficients

The association between scores generated from the TBP and parent and teacher SDQ total difficulties scores was explored using Pearson 'r' correlation coefficients (1-tailed tests). Two stages of exploratory analysis were completed. Firstly, correlations between TBP total positive and total negative summed scores and parent and teacher rated SDQ total difficulties scores were calculated, based on all 17 summed item codes. Secondly, correlations between measures were calculated based on the factor based subscale scores generated by Mueller (1996).

2.5.3 Reliability analyses of the TBP scores

Inter-rater reliability for the TBP total scores was calculated using Pearson 'r' correlation coefficients, for each of the 17 item codes. Mean correlations were found for both scoring systems i.e. total positive and negative scores, and factor based subscale scores. Internal consistency was tested by calculating alpha coefficients for total scores based on all items, and factor based subscale scores as described above.

2.5.4 Discriminant function analysis

Discriminant function analysis was used to explore the ability of the TBP measure to correctly classify children into different parent and teacher rated groupings. Grouping variables consisted of children rated as above and below clinical and borderline threshold on the parent and teacher SDQ; and parent and teacher ratings of concern.

3. Results

3.1 Sample characteristics

The final sample $n = 35$ consisted of 16 boys and 19 girls, and represented 50% of all children enrolled in the reception year at the school. All children were aged between 57 - 68 months (mean = 61.8). 14 of the children (40%) were in class 1 and 21 in class 2 (60%).

3.1.1 Non respondents

Two levels of selection were relevant to the inclusion criteria for the study (as described in the Method sections 2.2.1 - 2.2.2). Parents of 5 children who met the criteria for inclusion opted out of the TBP study. All were boys, 1 from class 1 and 4 from class 2, mean age 61 months. A between group comparison (1-way ANOVA) of parent and teacher SDQ total difficulties scores and global concern revealed that children who did not participate in the study did not differ significantly from children in the study group on parent and teacher SDQ total difficulties scores or ratings of concern. There is therefore no evidence of response bias in respect of the key study variables within the TBP non respondent group.

3.2 Parent and teacher ratings

3.2.1 Parent SDQ ratings

27 (77%) children fell into the 'normal' score range on the parent SDQ, 4 (11.5%) children fell into the 'borderline' score range, and 4 (11.5%) children fell into the 'abnormal' score range. A between group comparison (1-way ANOVA) of parent SDQ total scores revealed that boys and girls did not differ significantly in the study sample on the parent measure ($F = .02$, $df = 1$, $p = .90$). Table 4 over leaf summarises the results from the parent SDQ measure for the sample.

3.2.2 Teacher SDQ ratings

28 (80%) children fell into the 'normal' score range on the teacher rated SDQ, 3 (9%) children fell into the 'borderline' range, and 4 (11%) of children fell into the 'abnormal' score range. A between group comparison (1-way ANOVA) revealed significant differences between boys and girls in teacher SDQ ratings ($F = 4.24$, $df = 1$, $p < .05$), with boys showing higher mean total difficulties scores. Table 4 summarises the results from the teacher SDQ measure for the sample.

PAGE/PAGES
EXCLUDED
UNDER
INSTRUCTION
FROM
UNIVERSITY

boys showing a higher mean score ($F = 6.98$, $df = 1$, $p < .01$). Table 6 shows mean scores for boys and girls based on total scores generated using all item codes.

Table 6.

Distribution of TBP total scores (positive and negative themes) by sex (n = 35)

| TBP score | Boys | | | Girls | | | Total | | |
|-----------|-------|------|-------|-------|------|-------|-------|------|-------|
| | M | SD | range | M | SD | range | M | SD | range |
| positive | 15.38 | 3.95 | 8-23 | 15.11 | 3.76 | 5-21 | 15.23 | 3.79 | 5-23 |
| negative | 6.44 | 4.03 | 0-13 | 3.74 | 3.27 | 1-7 | 4.97 | 3.27 | 0-13 |

Note. Total scores derived from 17 item codes. Maximum positive total summed score = 54, maximum negative total summed score = 99.

3.3.1 Association between TBP total scores

Correlation coefficients were calculated to study the association between the total scores generated from the TBP using Pearson 'r' correlation coefficients (1-tailed tests). The correlation between TBP total positive score and total negative score was in the expected direction, but non significant (TBP total positive score with TBP total negative score, $r = -.27$, n.s).

3.3.1.1 TBP total scores and parent SDQ scores

Correlation coefficients were calculated to study the association between the total scores generated from the TBP, and parent and teacher rated SDQ total difficulties scores and subscale scores. The correlations between TBP total positive and total negative scores, and parent SDQ total difficulties scores revealed associations in the expected direction, but non significant (SDQ total difficulties score with TBP total positive, $r = -.11$, n.s; SDQ total difficulties score with TBP total negative, $r = .08$, n.s). Correlations between TBP total positive and total negative scores and each SDQ subscale score were all non significant.

3.3.1.2 TBP total scores and teacher SDQ scores

Correlations between TBP total positive and total negative scores and teacher SDQ total difficulties scores were in the expected direction, but as with the parent data, non significant (SDQ total difficulties score with TBP total positive score, $r = -.13$, n.s; SDQ total difficulties score with TBP total negative score, $r = .29$, n.s). Correlations between TBP total positive and negative scores and each teacher rated SDQ subscale score were all non significant.

3.3.1.3. Comparison of mean TBP scores across parent SDQ categories

Cut-off scores from the SDQ were used to divide children into two groups, those rated as above and below clinical threshold i.e. more and less likely to have emotional and behavioural difficulties. TBP total positive and negative scores were compared between groups to see if mean scores differed according to parent SDQ category rating. TBP total positive score averaged 14.25 (SD = 2.82) for children rated above clinical cut-off on the parent measure, and 15.52 (SD = 3.91) for children rated below clinical cut-off ($F = .32$, $df = 1$, $p = .58$). TBP total negative score averaged 5.50 (SD = 3.53) for children above cut-off, and 4.82 (SD = 3.31) for children below cut-off ($F = .03$, $df = 1$, $p = .86$). Both these findings were in the expected direction, but statistically non significant.

3.3.1.4 Comparison of mean TBP scores across teacher SDQ categories

TBP total positive score averaged 15.04 (SD = 2.70) for children rated above clinical cut-off on the teacher measure, and 15.29 (SD = 4.03) for children rated below cut-off ($F = .07$, $df = 1$, $p = .79$). TBP total negative score averaged 6.46 (SD = 3.66) for children above cut-off, and 4.64 (SD = 3.08) for children rated below cut-off ($F = .03$, $df = 1$, $p = .86$). Both these findings were in the expected direction, but statistically non significant.

3.3.1.5 TBP total scores and parent ratings of concern

Between group comparison (1-way ANOVA) was used to compare mean TBP total positive and negative scores with parent concern ratings. The analysis revealed no significant differences between TBP total positive score and TBP total negative score across the two parent concern groups (TBP total positive $F = 1.79$, $df = 1$, $p = .19$; TBP total negative $F = .84$, $df = 1$, $p = .37$).

3.3.1.6 TBP total scores and teacher ratings of concern

Between group comparison (1-way ANOVA) was used to compare mean TBP total positive and negative scores with teacher concern ratings. The analysis revealed no significant differences between TBP total positive score and TBP total negative score across the three teacher concern groups (TBP total positive $F = .55$, $df = 2$, $p = .58$; TBP total negative $F = 1.41$, $df = 2$, $p = .26$).

3.3.2 Reliability of the TBP

Two main measures of reliability in psychometric testing relate to the ability of different raters to produce the same results from the same material (inter-rater reliability), and the test's ability to demonstrate that the items which tap particular themes will generate similar responses (internal consistency). Pearson correlation was used to test agreement between independent coders on the TBP, and Cronbach's alpha was used to test the internal consistency of the total scores.

3.3.2.1 Inter-rater reliability

Inter-rater reliability was tested between two independent coders (LM and H-R) using a randomly selected set of ten videos from the sample. Each item code and total summed score for positive and negative themes were compared between coders, calculated as Pearson 'r' correlation coefficients. Mean inter-rater reliability across all 17 item codes was .75 (TBP mean total positive score, $r = .91$, $p < 0.01$; TBP mean total negative score, $r = .58$, $p < 0.05$; 1-tailed tests).

3.3.2.2 Internal consistency

Internal consistency was tested by calculating alpha coefficients for each total sum score (TBP total positive and TBP total negative). This analysis seeks to determine which common items are found consistently when repeated samples of variables are taken from a population of variables; and coefficient alpha represents the reliability of the score taken in a variety of situations (in this case, how consistently items contribute to total positive and negative scores). The higher the alpha coefficient, the more consistent the items which contribute to the total score. Whilst the expectation would be that the two total scores from the TBP are measuring different constructs, it is important to determine how consistent each is in measuring overall positive or negative themes to justify the use of summed total scores. The analysis revealed alphas of .42 for the positive items, and .59 for the negative items. This suggests that the consistency of the items which contribute to the separate total scores is low.

3.3.3 TBP factor based subscale scores (Mueller, 1996)

Mueller (1996) states that factor analysis of the TBP measure is “merely a tentative exploration of the relationships among the codes”, as the small sample size in his study ($n = 51$) substantially increased the subjectivity of the analysis and therefore the likelihood of error in the item loadings. On the basis of the low alpha coefficients for total positive and negative scores in this study, factor based subscale scores were generated, based on the 7 items extracted by Mueller (1996); as described in section 2.3.3 of the method.

Factor based subscale 1 (negative items) and factor based subscale 2 (positive items) mean scores for the sample were calculated as 2.54 ($SD = 2.32$) and 12.54 ($SD = 3.12$) respectively. Between group comparisons were calculated using 1-way ANOVA's to compare mean factor based subscale scores according to sex, parent and teacher rated SDQ groups, and parent and teacher rated concern category.

No significant differences were found between boys and girls scores on either factor based subscale (Factor 1: $F = .85$, $df = 1$, $p = .36$; Factor 2: $F = .08$, $df = 1$, $p = .78$).

No significant differences were found in factor based subscale scores between parent SDQ categories (Factor 1: $F = .13$, $df = 2$, $p = .88$; Factor 2: $F = 1.09$, $df = 2$, $p = .35$); teacher rated SDQ categories (Factor 1: $F = .21$, $df = 2$, $p = .81$; Factor 2: $F = 1.37$, $df = 2$, $p = .87$); parent rated concern category (Factor 1: $F = .96$, $df = 1$, $p = .33$; Factor 2: $F = 2.35$, $df = 1$, $p = .14$); and teacher rated concern category (Factor 1: $F = 1.47$, $df = 2$, $p = .25$; Factor 2: $F = 1.44$, $df = 2$, $p = .25$).

3.3.3.1 Internal consistency

Alpha coefficients were calculated for each factor based subscale score. The analysis revealed alphas of .45 for Factor based subscale 1 and .45 for Factor based subscale 2, which suggests that the internal consistency of the items within each factor based subscale score is slightly lower than found for the total positive and negative scores, based on all 17 item codes. It was therefore decided to continue the analyses using total summed scores based on all 17 item codes, as the factor based subscale scores were unstable in this study. This is, perhaps, unsurprising, as the Mueller (1996) factor analysis was based on a very small sample ($n = 51$), and excluded over half the items generated from the measure.

3.3.4 Can TBP scores accurately predict group membership?

Discriminant function analysis is a statistical technique used to predict group membership from a set of predictor variables (Tabachnick & Fidell, 1996). In this study, discriminant function analysis is used to determine whether TBP total positive score together with TBP total negative score can differentiate between children in the three parent/teacher groups, as defined by SDQ total difficulties score and ratings of concern. As discriminant function analysis requires a minimum of 20 subjects per predictor variable tested (Stevens, 1992), each analysis is performed separately rather than in combination in order to fulfil the assumptions required. It would have been interesting to run an analysis using all 17 item codes from the TBP, particularly as the internal consistency of the total summed scores is low. However, this would have required a sample of at least 340 children.

Using the direct entry method, the SDQ category (parent and teacher) and concern category (parent and teacher) served as the grouping variables, and the TBP total positive and TBP total negative scores were entered as the predictor (or classifying) variables in a series of separate discriminant function analyses.

3.3.4.1 Classifying children who are rated as 'cases'

Hypothesis 1: TBP total positive score and TBP total negative score will discriminate between children who are above and below clinical threshold on the parent and teacher SDQ.

Two direct discriminant function analyses were performed using TBP total scores as predictors of group membership, as defined by parent and teacher rated SDQ's. Parent and teacher SDQ scores were summarised into two groups - children whose total difficulties scores were above and below clinical threshold.

The first analysis was performed using parent rated SDQ category as the grouping variable, generating a canonical correlation of .11, chi-squared (2 df) = .37, $p = .83$. 31 (88.6%) cases were correctly classified into the specified groups. However, the 4 (3.4%) cases that were misclassified consisted of all children rated as above clinical threshold i.e. the TBP scores placed all cases in the below clinical threshold group; thus not discriminating between above and below cut-off on the parent rated SDQ.

The second analysis was performed in the same sequence, this time using teacher

rated SDQ category as the grouping variable. The canonical correlation was .22, chi-squared (2 df) = 1.57, $p = .46$. As with the parent rated groupings, TBP total scores misclassified all children who were identified as above clinical threshold in the teacher rated groupings.

3.3.4.2 Classifying children who are rated as 'at risk'

Hypothesis 2: TBP total positive score and TBP total negative score will discriminate between children who are within the borderline clinical range and those who are not on the parent and teacher SDQ .

In order to test the ability of the TBP total positive and total negative scores to discriminate between children rated in the borderline range, parent and teacher SDQ scores were summarised into two groups - children whose scores were above and below borderline clinical threshold.

As for hypothesis 1, two separate discriminant function analyses were performed. The first, using parent rated SDQ category as the grouping variable, generated a canonical correlation of .21, chi-squared (2 df) = 1.44, $p = .49$. The same pattern of classification was apparent, with 27 (77%) of cases correctly classified into the below borderline threshold group, and 8 (23%) children rated by parents as borderline misclassified into the below threshold group.

Teacher rated groupings generated a canonical correlation of .05, chi-squared (2 df) = .07, $p = .96$. All 7 (20%) children whose teachers rated them as above borderline cut off on the SDQ were misclassified into the below borderline group.

3.3.4.3 Classifying children who are rated as a cause for parent concern

Hypothesis 3: TBP total positive score and TBP total negative score will discriminate between levels of parent concern about their children.

Children in the sample were distributed into 2 of the 3 possible concern groups as no children were rated by their parents as a definite cause for concern. One discriminant function was therefore calculated, using parent rated concern category as the grouping variable, resulting in a canonical correlation of .25, chi-squared (2 df) = 2.03, $p = .36$. 27 (77.14%) children were correctly classified into the parent concern groups, including 1 child who was rated as causing some concern to their

parents.

3.3.4.4 Classifying children who are rated as a cause for teacher concern

Hypothesis 4: TBP total positive score and TBP total negative score will discriminate between levels of teacher concern about children.

Children in the sample were distributed across all 3 possible concern groups i.e. no concern, some concern, and definite concern. Using teacher rated concern category as the grouping variable, two discriminant functions were therefore calculated from the equation, resulting in a combined canonical correlation of .30, chi-squared (4 df) = 3.14, $p = .53$. 25 (71.43%) children were correctly classified into the teacher 'no concern' group. All children who were misclassified ($n=10$) were incorrectly placed in the no concern group (8 children who were rated as a cause for some concern, and 2 children definite concern).

4. Discussion

4.1 Overview of the section

In drawing together the findings from this study, and previous work based on the use of narrative measures, the discussion is organised into nine sections as follows:

- a summary of the study findings (Section 4.2)
- methodological issues (Section 4.3)
- clinical findings (Section 4.4)
- integration of the findings, including comparison of the TBP and MSSB narrative measures, and findings from previous studies (Section 4.5)
- implications for the measurement of child mental health (Section 4.6)
- theoretical implications (Section 4.7)
- therapeutic implications (Section 4.8)
- what next? (Section 4.9)
- conclusion (Section 4.10).

4.2 Summary of findings

The associations between TBP total positive and negative scores and parent and teacher total difficulties scores were all in the expected direction, but not statistically significant. The associations between mean scores derived from TBP total positive and negative scores and parent and teacher ratings of concern were also in the expected direction when groups were compared, although the results were not statistically significant.

Inter-rater reliability was high for TBP positive total score ($p < .001$), and less robust for TBP negative total score ($p < .05$). Alpha coefficients suggested that the internal consistency of the items which contributed to each total score were low. Factor based subscale scores derived from the original study (Mueller, 1996) suggested similarly low internal consistency.

Discriminant function analysis suggests that TBP total positive and negative scores do not reliably discriminate between children in the parent and teacher groups, due to a low correlation between TBP total theme scores and SDQ total scores. This indicates that TBP scores fail to identify children who are considered to be 'at risk' or above clinical threshold by parents and teachers. In this study, TBP total scores resulted in a high number of 'false negative' cases (poor sensitivity) which is

undesirable in a screening instrument (Boyle & Jones, 1985). There also appear to be limitations in the use of total negative theme scores as a clinical measure, due to the low inter-rater reliability. These findings suggest that further work on the psychometric properties of the measure with larger (clinical and non clinical) samples is required to improve the reliability and validity of the measure for use in clinical settings.

Scores from the TBP measure revealed significant differences between boys and girls in mean total negative scores.

4.2.1 Hypothesis testing

Discriminant function analysis suggests that the TBP's ability to discriminate between children in each of the parent and teacher rated groups was poor in this study. The F scores and canonical correlations, which indicate the degree of association between the variables were very low; suggesting that TBP total scores are unable to detect differences in children rated above and below clinically (and psychometrically) significant cut-off scores. The apparent high percentage of children correctly classified is no higher than would be expected by chance alone, and is only as high as it is because the largest majority of children fall into one group (below clinical and borderline cut-off on the parent and teacher rated SDQ; rated as a cause for no concern by parents and teachers). In many ways the finding is not surprising given the poor reliability of the measure. It therefore seems too early to expect analyses which are designed to test the validity of the measure to be statistically significant if reliability has yet to be established.

The findings did not therefore support the research hypotheses, as TBP total positive and negative scores were unable to:

Hypothesis 1: identify parent and teacher rated 'cases'.

Hypothesis 2: identify children rated by parents and teachers as 'at risk'.

Hypothesis 3: discriminate between children in parent rated concern groups.

Hypothesis 4: discriminate between children in teacher rated concern groups.

4.3 Methodological issues

Several methodological issues are likely to have contributed to the poor internal consistency of the measure. These include the small sample size, the administration

one (an overall total of thirteen parents opted out of the TBP study, five of whom completed the parent questionnaire). Sixty nine teacher SDQ's were completed, therefore more inclusive criteria would have resulted in a total TBP sample of n=56 (n=35 parent and teacher measures, n=21 teacher only measures).

Three children's TBP data were lost through video recording errors. This was unfortunate, and only really avoidable through reducing time constraints and increasing support to the testers e.g. by collecting data in pairs. This was not possible within the constraints of the study.

4.3.2 Administering the TBP

Several difficulties became apparent during the administration and scoring of the TBP, which are likely to have had a direct impact on reliability.

4.3.2.1 Preparation for testing

Testers practised delivery of the story stems with H-R, each other, and children from their own families in order to learn the stories verbatim. Agreement was reached in how long to continue prompting children to tell a more detailed story, and to describe the feelings and actions of the bears. The Directions for Administration (see Appendix 7) emphasises the acceptability of saying "go on", "what happens next?", "how does it turn out in the end?"; whilst avoiding any questions which are leading (e.g. "is your bear thinking he's done the right thing?", "did he feel angry?"). In practice, it remained difficult for testers to always remember the precise wording of each story stem, which led to the need to refer to the directions on occasion during administration. This improved over time, but influenced the 'playfulness' of the interaction for some children, as discussed below.

4.3.2.2 'Playful scenario'

Mueller (1996) likens the administration of the TBP to a 'playful interaction', where the stories are designed to elicit a 'playful response' from the child. The tester should not, therefore, present the story stem material in a formal way, but rather prompt the child to enjoy the activity, and enter fully into the imaginative world of the bear family. Aiming to achieve a playful frame of reference is understandable and desirable if opportunity is to be maximised for the child to project him/herself fully into the scenarios, and develop detailed constructions from the prompted hypothetical social and moral dilemmas. Further practice in learning the stories verbatim was required.

In practice, both children and testers varied in their abilities to interact playfully with a relative stranger. Testers appeared to experience a form of 'test anxiety' where they became concerned about their performance e.g. were they administering the measure correctly, were they pushing the child too hard or too little, or over/under prompting for feelings; and would they make a fool of themselves on camera? Children were often hesitant initially, and some required longer than others to settle into the exercise. The introductory 'honey pot' story was administered to all children as a warm up exercise for both parties. In this study, several children repeated themes from this story in the first one or two test stories, which suggested variability in the amount of time required for practice. Mueller (1996) states that any repetition of themes from earlier stories always generates a code of 'immaturity', yet there were qualitative differences between children who eased slowly into the stories and repeated themes in a manner which seemed to relate to 'checking out the process' with the tester, and those children who returned to specific content themes on a frequent basis.

Generating playfulness was somewhat constrained in this study by the characteristics of the setting in which the TBP was administered. Privacy was non-existent, and although the children were familiar with using the research area for individual work, there were regular interruptions, and at times even the most focused children were distracted by other people/activities which were taking place in the corridor. Time constraints were also significant. Playtimes and lunchtimes could not be timetabled for TBP administration as the area was a busy through route for staff and children throughout these periods. As only one tester could collect data at any one time, this put considerable pressure on the researchers to administer the measure to as many children as possible during each visit to the school in order to minimise time delay between parent, teacher and child measures. Tester tiredness may therefore have been influential.

4.3.2.3 Prompts for feelings

Testers consistently asked children how the bears were feeling in response to a specific event or action within each story, but did not always consistently follow this up with further prompts to continue the playful scenario. Many children responded to the questions about feelings with one word answers (e.g. fine, happy, sad). This was problematic in scoring, as described in section 4.3.3.2.3 below. Differences were apparent between the testers in the quality of probing. Despite considerable training,

one tester remained less able to encourage the children to expand on their narratives. This is likely to have been related to her non clinical background.

4.3.2.4 Remaining within the 'fantasy frame'

A general difficulty in administering the measure related to testers not always remaining within the 'fantasy frame'. On several occasions, both testers referred to "you" (e.g. "how did you feel?") rather than remaining within the 'third person' and referring to the 'special bear'. Several children appeared not to notice, but some children did respond in accordance with the question i.e. in the 'first person'. This had implications for scoring, as the Scoring Directions clearly state that any loss of identification with the bears in the form of taking on a bear role renders the story 'immature'.

4.3.2.5 Video-recording the TBP

Video-recording the TBP had several advantages in terms of coding, but also some disadvantages. The quality of the equipment, particularly in relation to sound recording, contributed to difficulties in hearing clearly the precise words used by the children. This was mainly due to background noise taking place in the area of administration. Placing the camera behind the children mitigated against using non verbal expression and gesture to aid the coding process. This is acceptable within the criteria for scoring, in that non verbal communication is not scored in it's own right (reference is made to audio recording in the directions for administration); but seeing a child's non verbal response significantly improves clarity for coding. The presence of the video camera also acted as a magnet to passing children. This was extremely difficult for the child carrying out the task, who would inevitably be distracted; and at times, both the child and tester would lose the thread of the story.

4.3.3 Scoring the TBP: Interpreting the scoring directions

LM and H-R conferred to prepare for coding in order to agree interpretation of the Scoring Directions (Mueller, 1996), and their application of the rules to the narrative material presented by the children. Several difficulties became apparent during this process.

4.3.3.1 General scoring rules

Several general rules are described in the scoring directions (Mueller, 1996), all of which apply to all stories presented. Some rules were difficult to understand

theoretically, and others were clear but difficult to apply consistently in practice.

4.3.3.1.1 Coding what is observable

Stringent application of this rule was applied throughout. Notes were kept to record information that was ambiguous, or felt to be relevant but uncodable within the scoring criteria. This rule, whilst clear and understandable, contributed to the low frequency of coded themes throughout (maximum possible range for each total theme score across 9 stories is 0 - 54 for positive total score, 0 - 99 for negative total score. Actual range in this study was 5 - 23 for total positive score, 0 - 13 for total negative score).

4.3.3.1.2 Coding only the first occurrence of each code

All codes were used only once per story stem. Again, notes were kept, and coders tried to indicate each occurrence of each theme on a separate sheet in order to compare results from 'first only' coding with 'all coding'. This became unreliable and was therefore abandoned. It was generally felt that the number and frequency of themes present in each story would be a useful measure, as several children appeared similar on the basis of total scores, yet were very different in their responses to the story stems. As for the first general rule, coding the first occurrence only of each theme significantly reduced the total number of codes marked as present for each child.

4.3.3.1.3 When 2 codes fit use the first only in a given order

This rule was difficult to understand and apply. The 4 codes which the rule applies to include sexual/oral drive expression, aggression, main character vulnerability, and main character devalues others. For example, children who express a sexually relevant theme (look up mother bear's dress; undress the bears and hit them) could not be coded as aggressive, vulnerable or devaluing in relation to another character at the same time, even though intuitively, these codes could be both directly relevant to the context of the theme and informative about the child's internal representations. Thematic content together with qualitative expression (e.g. aggressive or vulnerable) seems to be at the heart of understanding the enactments of troubled children, yet this rule rejects the possibility of coding two themes consecutively. A more sensible solution might be to be over inclusive in thematic coding in response to clinically significant themes rather than the reverse.

4.3.3.1.4 Main character switches between characters

This rule appeared simple but was tricky to apply, because there were often times within the children's stories where two characters were the focus of the narrative, rather than one more than the other. The general agreed rule of thumb between the coders was that all characters are main characters when coding their actions and interactions, whilst the child's 'special bear' remained the primary main character throughout.

4.3.3.1.5 Themes suggested in the story stem are not coded

This rule was clear but not always straightforward to apply. Some children would return to the theme suggested in the story stem more frequently than others in constructing their narrative response, whilst others would not mention it again. Some children also took longer to settle into the exercise, and, as described in section 4.3.2.2 above; referred to the warm-up and first stories in checking out that they were doing the right thing. Themes which were relevant were sometimes used in this context, yet were not coded, as the rule was adhered to throughout.

4.3.3.1.6 Imaginary characters are coded in the same way as other characters

This rule was clear to understand and apply, although it did not occur frequently during the administration.

4.3.3.1.7 Scoring behaviour between the subject him/herself and the bears

Any action or behaviour initiated by the child themselves towards the bears i.e. outside the fantasy frame always generates an 'immature story' code. Any "I" response is also considered to be indicative of an immature story, as the child has lost their pretend role within the imaginative scenario. As described in section 4.3.2.4 above, testers themselves sometimes lost the fantasy frame before the child, and were noted on occasion to prompt a child into responding in the 'first person'. Agreement was reached between the coders that the story would not receive an 'immature story' code if the tester had initiated the response. This may have contributed to the low frequency of 'immature story' codes.

4.3.3.2 Scoring the TBP: Using specific item codes

A small number of individual item codes were problematic in their application to children's stories. These are discussed below.

4.3.3.2.1 Statements of relationship are scored twice

This rule applies to the valuing and devaluing codes (main character to and from others). The rule is clear in its meaning i.e. “we love each other”, “they played nicely together” generates reciprocal codes of ‘main character valuing’ and ‘main character being valued’ at the same time. Applying the code was more difficult, as if a code had already been scored once, additional occurrences of that code, including mutually valuing/devaluing relationship codes could not be scored. This important and qualitatively different information was therefore frequently lost. Even when mutually occurring themes could be scored (usually near the beginning of a story stem), the scoring record sheet did not easily lend itself to recording which codes were derived from single events, and which from mutually occurring events.

4.3.3.2.2 Valuing/devaluing codes

These codes were the most frequently used for all children. However, although conferring had led to agreement between LM and H-R in how to apply the codes, consistency was very difficult to maintain, particularly for the ‘devaluing others’ code. The scoring directions clearly recognise the lack of absolute clarity for all responses given, as what is acceptable in one cultural setting may not be considered acceptable in another. For example, in response to Story 4: The soup bowl story, children described a range of parental discipline styles e.g. the bear was “shouted at for being naughty”, “smacked and sent to bed”, “shouted at for being naughty and grounded”; some to the degree of generally being considered inappropriate, some acceptable, and others somewhere between the two extremes. The subtle (and sometimes not so subtle) nuances of difference between codes relate to a wide range of factors. These include the child’s own frame of reference, the tester’s exploration of (in this particular example) the discipline style and its effect on the bear’s feelings; and in relation to scoring, the coder’s own experience and view on what is considered to be appropriate and inappropriate discipline for the behaviour. Slightly different interpretations of such examples have markedly different impacts on the final score. Acts of discipline which are considered appropriate generate positive ‘valuing of main character’ codes, and an absence of ‘devaluing main character’ codes; based on the premise that appropriate limit setting and discipline are nurturing and caring parental acts. Therefore, a different interpretation of a single act within each stem will not only lead to a higher or lower score on one code, but a simultaneous increase or decrease in an opposite pole code. These factors led to some inconsistency, particularly in the use of the ‘devaluing’ codes. This is born out

in the reliability analysis.

4.3.3.2.3 Coding emotional expression

Children were frequently and consistently prompted by testers to describe how characters were feeling, as instructed in the administration directions. However, in coding positive and negative affect (in main characters and others), scores are generated only in response to spontaneously expressed emotion (or elaborate expression of emotion if prompted e.g. “he was really horrid and made her cry a lot”). Most children responded to prompts with one or two word answers, therefore a significant number of affect responses were not included. This contributed to the low frequency of affect codes within the sample.

4.3.3.2.4 Summary codes

Three summary codes are used in response to each whole or part story, ‘story disoriented’ (things happen in the story which do not make sense, or are contextually strange); ‘immature story’ (underdeveloped story telling ability, reuses themes from previous story stems, loses the story fantasy frame and takes on the main character role); and ‘child distracted/impulsive’ (child frequently wanders off the task). ‘Story disoriented’ codes are included in the total scores, whilst the remaining two summary codes are not.

‘Story disoriented’ was the most frequently used of the summary codes, in response to children’s stories wandering and losing context. Although it is common for 5 year old children to engage in self talk during play, which can appear to be unrelated to the presented story, it was clear when it was appropriate to apply this code.

Children’s narratives were qualitatively different when they had lost the point of the story as compared to children who elaborately described the actions of the bear’s in response to the dilemma in a process of self talk which may not have been immediately understandable by the tester.

‘Immature story’ was infrequently scored, as most children were able to construct coherent narratives and remain within the fantasy frame.

The testing environment distracted children for the reasons described in section 4.3.2.2 above. Agreement was reached between LM and H-R that where children’s attention wandered as a direct result of interruption, the ‘child distracted/impulsive’

code would not be scored.

4.3.4 The use of composite total scores

4.3.4.1 How appropriate is it to add together total scores?

It appears inconsistent to elicit information from children using story stems which are specifically designed to tap different themes, then combine the scores into summed totals. It would make more psychometric (and clinical) sense to sum the total scores from themes generated within each story, then measure the reliability of each separate story narrative, which is similar to the procedure applied in several studies which use the MSSB. This study demonstrated poor reliability from using total scores. Whilst this could also be related to inconsistencies in administration and scoring across testers and coders, it is likely that the total scores in themselves are unreliable. An alternative analysis could treat each story narrative as a separate variable in its own right. In this study, each child generated 9 complete story narratives, using a maximum possible 17 coded items in each. A total of 315 narratives could then be analysed using item and total scores, and factor analysis on a sample of this size may begin to define groups of items which are more reliable in constituting factor based subscale scores.

4.3.4.2 Items which do not contribute to total scores

Two of the three summary codes are not included in the scoring process which may be psychologically important. In this study, a small number of children were observed to be distracted/impulsive in some stories ($n=7$), and immature in their story telling abilities ($n=2$), over and above the influences imposed by the environment. Interestingly, 5 of the 7 children who scored on the distracted/impulsive code were also rated as having difficulties by parents or teachers (2 were highlighted by both informants); and both children who scored on the immature code were also highlighted as having problems (1 by their parent and 1 by the teacher). These findings may suggest that the codes which are not included at any stage of the generation of total scores are important in highlighting children's difficulties.

4.3.5 Factor analysis (Mueller, 1996)

4.3.5.1 Items selected for the factor analysis

10 items were selected in the Mueller study to enter into a factor analysis (see method section 2.3.3). The 7 items which were not selected comprised:

- sexual/oral expression
- physical punishment
- resourcefulness in others
- positive affect
- negative affect
- positive affect in others
- negative affect in others

It is unclear why all themes associated with affect are not considered in the analysis, nor the infrequently used but clinically significant sexual/oral expression and physical punishment. Several items from the total 17 are paired (feature in main character and others). Main character resourcefulness is included in the factor analysis, yet resourcefulness in others is not, which also appears inconsistent.

Items loaded into 3 factors in the factor analysis (see method section 2.3.1.3), following which 3 items were excluded. These items consisted of devaluing from others, main character vulnerability, and sad story ending. Devaluing from others was excluded on the basis of a clear criterion. However, the remaining 2 items did not load consistently with items of a similar theme (e.g. sad story ending was positively correlated with positive items), but did meet the defined loading requirement. In addition, a high number of emotionally driven items which were also excluded either before or following the analysis. These appear to be the very items which may tap internal representations which are not only directly relevant to children's emotional and behavioural status, but are also less visible to those around them. Mueller (1996) provides no explanation of why all these items were excluded, which raises questions about the validity of the analysis and conclusions drawn.

4.3.5.2 Replicability of Mueller (1996)

The lack of validity of factor analysis with small samples raises questions about the results obtained in the Mueller (1996) study, and suggests that replication of the findings is unlikely. Comrey & Lee (1992) suggest, as a guide to factor analysis, that sample sizes of 50 are very poor, 100 poor, 200 fair, and as a general rule, "it is comforting to have at least 300 cases for factor analysis". Lack of clarity regarding the scoring directions for some items, coding bias in relation to variability in what is considered culturally acceptable, and the mysterious disappearance of 5 emotionally based items, all raise several important questions about the reliability of the factor

based subscale scores. In this study, the low alpha coefficients for each of the factor based subscale scores bear this out.

4.3.6 Improving reliability

The internal consistency of the measure needs to be far greater if the TBP is to be considered psychometrically sound. This means that codes drawn from the stories need to increase in accuracy and dependability in tapping the same constructs in the same way. Similarly, each individual item and (where relevant) composite scores need further clarity of definition for coding purposes to increase inter-rater reliability. Several possible routes to improving internal consistency and inter-rater reliability are discussed in section 4.9 below.

4.4 Clinical findings from the TBP

Children's responses to the story stems elicited a wide variety of clinically salient responses. Emotional vocabulary tended to be limited to three or four words for the majority of children (happy, sad, cries, angry); but actions and story content varied considerably. Examples of clinically relevant responses are described below.

4.4.1 Positive themes

Children who were focused and engaged in the fantasy world of the bear family generally told more detailed and descriptive stories, and generated higher positive total scores. Representations of rule adherence and delayed gratification were more common in these children. Story 3: The blackberry bush stem (main character asks mum to play football) frequently elicited prosocial, regulated responses, where main characters waited for mother bear and played alone, without distress; later returning to observe mother's actions. On five occasions, main character bears solved the problem by seeking out their sibling to join in instead.

Resourcefulness and positive affect codes were more frequently used with this group. Some main character bears created ingenious solutions to rescuing their sibling in the Story 5: The rock story e.g. one child described rolling the injured sibling all the way back to parents; another built a shelter from leaves and stones to protect the sibling while main character bear ran to find a parent. Greetings were also considerably warmer than for the group who demonstrated a higher frequency of negative themes e.g. in Story 9: The return, main character bears ran to greet parent figures from the fishing trip, keen to discover if any fish had been caught;

hugged then waved to mother figure on arriving at school in Story 6: Bye, then ran out to show her their school work at the end of the day, to be greeted by mother bear with a hug and some sweets. These children more often were able to solve the problem in Story 7: The problem (main character bear's turn to answer the question in school and write the answer on the board).

4.4.2 Negative themes

4.4.2.1 Themes associated with 'externalising' problems

Several children described problem solving strategies and discipline themes in their stories which contained several negative action and emotional themes. Story 1: Wagon (sibling takes toy wagon) elicited aggressive responses from some main character bears in their attempts to regain control of the toy, and a small number of bears would fight if the sibling was resistant in any way. Sibling representation was also evident in Story 8: Cave (parent bears leave children in cave while they go fishing). Main characters who left siblings to cry, or competed over toys in the absence of parents suggested a lack of prosocial response.

One main character bear expressed considerable aggression towards mother bear in several story stems. In Story 3: Ball (main character bear wants mother bear to play but she is busy picking blackberries), the child's main character bear responded to mother bear's inability to play because she was picking blackberries by repeatedly throwing the ball at mother bear's head, and laughing almost hysterically each time the ball hit her. This child also represented several themes associated with ineffective behaviour management. In Story 4: The soup bowl, the main character bear knocked the soup over, laughed for a few seconds as the stem was delivered, then ate the parents soup. No discipline was applied.

The physical punishment and aggression items were infrequently coded, but appeared to offer insight into children's experiences. One child's mother bear hit out frequently to get father bear and main character bear to do as she wanted within the context of the story. Physical punishment was also used to chastise main character bear for knocking over the soup in Story 4.

4.4.2.2 Themes associated with 'internalising' problems

Main character devaluing and devaluing from others were difficult to code consistently, but were apparent in several children's story stems. Parent figures

were noted to respond to main character requests with phrases such as “not now ... bear”, “don’t be silly”, “shut up”; or main character bears would ignore instructions and requests from parent figure bears by walking away or playing with a toy. Sexual/oral expression was only coded once for one child in the study, who, interestingly also scored on codes representing aggression, main character vulnerability, story disoriented and distracted/impulsive. This child was rated as falling in the borderline clinical range on the parent rated SDQ, clinical range on the teacher rated SDQ, and was a cause for definite concern to his teacher. This child’s total positive and total negative TBP theme scores were low (total positive = 11, total negative = 8) suggesting that in this example, individual item codes were considerably more meaningful as indicators of difficulty.

Items which represented negative affect in main character and others items were coded relatively infrequently in comparison to the total number of feelings described, as most expressions of affect were prompted by the testers. However, the presence of main character negative affect in 5 children coincided with expressions of difficulty by parents and teachers (SDQ and concern ratings). Again, this is clinically relevant, particularly as affective themes are likely to be associated with emotional symptoms. Two children scored on the main character negative affect code in the absence of parent or teacher rated difficulties. Perhaps these are the very children the test is able to identify at a stage before problems have been noticed by close adults.

Stories with sad endings showed a similar pattern i.e. infrequent coding, but associated with parent or teacher rated difficulties. Children who told sad stories who were not highlighted by parents or teachers may also be expressing internalised emotional difficulties which are not apparent to those around them.

4.4.3 Sex differences

The TBP highlighted significant differences between boys and girls in their expression of thematic material. In this study, differences were apparent in the area of total negative theme scores, with boys showing a higher mean score in comparison to girls. This finding is promising as most early onset child mental health problems are more common in boys, including those associated with developmental process (e.g. autism, speech and language disorders); and conduct related problems (Taylor et al., 1986). Physical and psychological vulnerability is also greater in boys (Schaffer, 1990), with a higher incidence of disorder (Richman et al;

1982) and greater susceptibility to adverse reaction in the face of adversity (Zaslow & Hayes, 1986).

The narrative basis and current scoring criteria of the TBP therefore appear able to distinguish between boys and girls in their expression of psychologically negative themes. This suggests that with increased reliability, the measure has the potential to identify individual differences between boys and girls.

4.4.4 Statistical vs clinical significance

The stories told by children in the study support findings from previous UK and US based studies which highlight the clinical significance of the use of narrative story stems with young children. If reliability increases (internal consistency, inter-rater reliability and stability), validity is also likely to increase. The story stems generate themes which tap a range of psychological domains which are relevant to children's development, and to mechanisms of risk and resilience (e.g. inconsistent parenting style, conflict, aggression; adaptive social and emotional development, positive self esteem, prosocial behaviour). Although these themes are not collected through the use of an entirely systematic method, further clarity of coding will increase the understanding of the meaning attached to specific scores for specific items and children. The current clinical significance associated with the measure strongly suggests that it will be valuable to work towards improving the psychometric properties of the measure.

4.5 Integration of findings: Comparison of measures and links with previous US & UK studies

Factors which contribute to child mental health problems were apparent in the children's narratives. Aggression, family conflict, punitive discipline style and parent rejection were all observed in the study, and on two occasions, led to the initiation of intervention. Children's stories led to opportunities to observe their enactment of main character emotion regulation, a process which would not be open to direct interview style assessment, yet was illustrated with ease during the playful narrative interactions. This supports the view that children in the pre and early school stage are comfortable with this mode of play to express their internal representations and that this may be the most useful medium through which to increase our understanding of the child's perspective directly from the pre/early school child.

4.5.1 Comparison of TBP and MSSB

The TBP and MSSB clearly have several things in common. They both elicit internally represented constructs from children using incomplete story stems, designed to tap specific themes, by posing hypothetical social and moral situations and dilemmas. The methods through which the stories are delivered in both measures overlap considerably, with the use of age appropriate toys and props to engage children into the fantasy world of play. The very broad principles of coding the material generated from the stories are also similar, where both seek to code actions and interactions which are observable from children's verbal and non verbal responses through the use of video or audio recordings. However, specific rules for coding the MSSB are considerably more detailed than those available for the TBP.

4.5.1.1 Scoring the MSSB

Codes are divided into separate phases of delivery, and the length of time taken for children to complete each phase is noted. Content themes are scored as present or absent in each story using tally marks on a summary coding sheet. Emotional themes and performance themes are coded using rating scales according to the presence or absence of the theme, and if present; the degree of response represented in the child's story. Additional codes relating to the child's general behaviour and reactions are tied to particular time points within the delivery and response to each stem. These include scales to code startle reaction (i.e. to record the absence, presence, and degree of involuntary physiological reaction at any point), initial response to the story problem (e.g. aggression/destruction/injury, denial/no explanation, denial/explanation, avoidance of conflict, help seeking, help arrives, child resolution, self blame to relieve conflict, fear/anxiety response, positive content, angry/negative content, story-stem response, repetition); final content (e.g. positive, neutral, negative, fear, angry, aggressive); ending transformation (to neutral or positive/negative affect); emotional incoherence (e.g. lack of clarity, shift to negative theme which does not make sense); emotional shift (coherent); incongruent affect in response to a negative situation (e.g. smiling when faced with a frightening situation); reaction to blocked goals/desires (e.g. response when faced with situation where they cannot have what they want); and child representations (e.g. counter dependence/child power).

Content themes generate discrete data (i.e. the theme is present or absent), whilst emotional and performance themes generate continuous data based on ratings of

the child's response. Studies vary in their approach to the use of composite scores, which are often detailed in their collation, but the overall scoring process retains all information generated from children in the form of codes to represent all the thematic material within each story.

4.5.1.2 Scoring the TBP

In contrast, and in some ways to its advantage, the TBP is far simpler to code. There are no phases of coding nor associated 'times taken' to record for each phase, and no rating scales to allow variations in children's responses to be captured. The first occurrence only of any code is scored within each story. Scores are derived from the presence of a specific theme, and entered onto a summary coding sheet. A score of 0 or 1 is possible for each of the 17 coded themes in each story, resulting in a total score ranging from 0 - 9 for each theme. Total positive and negative scores are calculated at the end.

Perhaps the most significant differences in the scoring criteria for the measures, in light of the poor internal consistency found in this study are the coding of the first only occurrence of each theme in the TBP in contrast to every occurrence in the MSSB; and the use of rating scales in addition to dichotomous present/absent codings in the MSSB.

4.5.2 Findings from previous studies

A growing literature from the US and UK over the last ten years supports the usefulness of narrative story stem techniques with young children. Several issues are relevant to findings from this study, both in understanding the low reliability of the TBP, and in planning further research to address this problem.

4.5.2.1 Administration and coding

The administration and coding of material is central to the reliability and subsequent validity of the measure. Kochanska et al (1996) describe their refinement of both aspects of the MSSB, which led to the use of standardised questions at the administration stage. This significantly aided the coding process, and whilst a potential disadvantage may relate to a move towards a more 'interviewing' style with children (which is inherent with difficulties associated with poor reliability), the study suggests that the key to this is specifically to ask children to adopt the role of the protagonist in answering questions about each dilemma posed in each story. This is

similar to the main character within the TBP, which for most children in this study (and in the previously reported study using the TBP) is their 'special bear'. Asking children what their 'special bear' was feeling/thinking/doing through the use of standardised prompting at the end of each story would increase the likelihood of:

- reducing the difficulties associated with not coding prompted responses
- extending the completion of the story with children whose endings were vague
- returning to the story ending in the event of distraction
- systematic and consistent scoring of each occurrence of each theme, and
- minimising total variability in coding.

Internal consistency and inter-rater reliability are reported as high in several US studies. This is likely to relate to the different types of coding used in the MSSB for different types of data i.e. discrete content themes are counted (then aggregated across the whole test), and rating scale scores are converted into mean scores for each emotion and performance theme. Whilst similar difficulties are likely to arise in interpreting thematic material consistently within and across coders in any coding system, the opportunity to measure qualitative differences in children's descriptions of themes carries a high level of face validity; providing the definitions for applying the specific anchors in each scale are clear and precise. In contrast, the TBP coding does not attempt to capture this level of detail.

4.5.2.2 The use of single stories: The 'mix and match' approach

The use and development of specific stories for the exploration of specific themes is a common feature in studies using the MSSB. Stories from the TBP are also designed to tap specific themes, but there is considerable overlap between stems across the test e.g. separation/reunion in two stories, sibling relationship quality in two stories, parental discipline style in three stories. This suggests that each story does not focus predominantly on any one specific theme, but rather a mixture of themes salient to the hypothetical dilemma posed; so the use of single stories may not be viable. However, it may be useful to increase the specific nature of each story stem within the TBP, which would suggest that particular themes would be represented more frequently in specific stories. Stories in their own right could then be analysed for reliability.

4.5.2.3 General codes

Mueller and Tingley (1990) describe the development of a general "quality of family relationships" code from early findings. However, general codes are not used in the later TBP to summarise thematic content, rather only to capture the child's behaviour (immature, distracted/impulsive), and the ability to tell a coherent story (story disoriented). General codes for each story could be useful e.g. overall, what was the quality of family relationships/prosocial behaviour/parental responsiveness in the story told? Based on finding from studies carried out in the US, rating scales could provide anchors, which, in combination with frequency counts of each theme would make use of considerably more of the material presented by children; and may enhance the reliability and validity of the measure.

4.5.2.4 A new measure based on the TBP: The MUG (Mueller, in press)

Recent development of the TBP in the UK has led to the construction of a narrative measure for use with children in middle childhood, aged 7-12 years (MUG: Mueller-Ginsberg Story Technique, Mueller; in press). Administration is based on the use of specially designed picture cards which are used to prompt narrative material from children. The measure is similar in format to the Thematic Apperception Test (TAT: Murray, 1943), and the equivalent children's version, the Children's Apperception Test (CAT: Bellack & Bellack, 1974); but is derived from a different theoretical standpoint. The coding of children's responses to the MUG is based on the same system which is used for the TBP, with additional items designed to tap themes which are more relevant to older children (e.g. death wishes, thoughts of suicide, motivation for achievement, boredom, frustration). As with the TBP, each theme is scored only once during each story, and total positive and negative theme scores are summed across all stories.

The MUG has been administered with a much larger sample than was reported for the TBP ($n = >600$ referred and 'matched' non referred children). The same factor analysis of the results, as applied in the study using the TBP (Mueller, 1996) has led to a six factor structure, consisting of three negative and three positive factors. Interestingly, and of particular relevance to findings from the current study, is the pattern of clustering which occurs in factor 1 (sexual/oral drive expression, impulsive/distracted, immature, and story disoriented). Although caution is necessary in making direct comparisons between the new Mueller (in press) study and this study, it is important that three of the four items which group together in the

first 'MUG' factor are items which do not contribute to the total (or factor based subscale) scores in the TBP. This is encouraging, based on the apparent clinical associations between a small number of children's scores on these items, and parent and teacher rated emotional and behavioural problems in this study sample.

4.6 Implications for the measurement of child mental health

Several studies have found that in child psychology and psychiatry data from multiple informants is necessary if problems are to be effectively identified. When two or more informants respond in a similar way to questions which seek information about a third person, then it is likely that their aggregated responses will correctly identify the presence or absence of a problem (although not necessarily the nature of the problem) (Bird, Gould & Staghezza, 1992). However, when informants provide discrepant information, aggregate responses are not so useful in highlighting the presence or absence of a problem.

Where information is available from multiple informants about child mental health, parents and teachers are generally better informants of 'externalising' problems, whereas children themselves are better informants of 'internalising' symptoms (Edelbrock et al., 1996; Herjanic & Reich, 1982). Studies which have included informant data from children have mainly focused on children in middle childhood and adolescence, given the lack of reliable instruments available for use with younger children. However, Kolko & Kazdin (1993) examined agreement between child, parent and teacher reports of emotional and behavioural problems in a sample of community and clinic referred children in the 6-13 age group. Results revealed low to moderate correlations between informants, with parent-teacher correlations generally highest, and child-parent correlations the lowest. The level of agreement between all informants was significantly higher for 'externalising' problems.

Offord et al (1996) in a study using a large general population sample (n=1,134) conclude that assessment data from multiple informants should not be aggregated, but considered separately, as each informant may see objectively different problems according to the setting they are from, and their own perception of what constitutes a 'problem'. This study was based on parent and teacher ratings of child mental health, but it is logical, based on the evidence that children report more problems overall than parents and teachers (Stanger & Lewis, 1993), and are significantly more likely to report 'internalising' symptoms; to extend this view to including

separate information from children within multiple informant data.

Results from the study suggest that the technique is less applicable as a whole population 'screening' instrument, but rather focuses on a much needed method for the direct assessment of children on an individual basis, to identify themes of relevance which may indicate disrupted attachment, emotion dysregulation, and emotional and behavioural difficulties. The development of the TBP has the potential to make a considerable contribution to the assessment of children in the preschool age group, by providing young children with the opportunity to act as their own informants; perhaps at a stage when discrepant information has been gathered from adult informants.

4.7 Theoretical implications

The TBP embraces four main theoretical perspectives, and is described by Mueller (1996) as resting within a contemporary cognitive approach (see Introduction section 1.5.4.1). Each perspective recognises that children in early childhood have limited direct access to their internally organised thoughts and feelings, and that 'imaginative pretence' and 'imaginative reality' (Emde et al., 1997) are most effectively tapped through the medium of play. During playful interaction, children let down their guard, and reveal elements about their construction of self and other unintentionally (Mueller, 1996).

Several studies which have used narrative measures with children have focused on understanding the association between attachment relationships, emotion regulation and parent and teacher rated emotional and behaviour problems. The assessment of attachment relationships is highly complex, and must take into account how an individual organises their behaviour and regulates emotion (Sroufe, 1994).

Ainsworth, Blehar, Waters & Wall (1978) developed an 'in-vivo' observational procedure to assess quality of attachment in infants. The system consists of 8 episodes or contrived situations which lead to increasing amounts of stress for the infant in experiencing separation and reunion with the caregiver. Whilst infant reaction to separation is important in tapping the quality of the relationship, it is the quality of the reunion which has more validity in assessing the capacity of the dyad to manage the stress of the separation, and hence offers a reliable perspective on the quality of the attachment relationship (Field, 1985; Schore, 1994). The Ainsworth et al., (1978) system is based on the use of four categorical scales which are

designed to capture the effectiveness of the reunion between infant and caregiver. Each scale provides behavioural anchors, and the coder's task is to find the best match between what is observed during each episode and the descriptive anchors. The ratings are then grouped together to classify the overall pattern of the attachment relationship according to three main categories (secure, anxious/resistant, anxious/avoidant).

Themes from the TBP and MSSB are directly linked to the ratings of behaviour in younger children which are used to classify attachment patterns e.g. valuing, nurturing, response to separation and reunion, rule adherence, empathy. The stories are designed to identify themes within children's narratives which directly relate to the internal organisation of behaviour and associated emotion regulation. Following the development of more reliable administration and coding procedures, narrative measures have the potential to contribute to increasing understanding of attachment patterns in preschool children, which in turn will increase theoretical understanding of the processes involved in the onset and maintenance of emotional and behavioural problems; and the developmental pathways which interact to increase or decrease the likelihood of such problems arising in the period between infancy and middle childhood.

4.8 Therapeutic implications

Clinical findings from the study suggest that the TBP may also be a useful therapeutic tool. In assessment, it has the potential to be used with children in highlighting themes for clinical focus and change. In therapy, it may provide an appropriate medium through which to act out salient problematic situations, and test out alternative solutions to problems through providing children with an age appropriate, meaningful context to construct different stories to resolve their own dilemmas.

The use of the TBP as a therapeutic tool can be linked directly to the framework of narrative therapy. Narrative psychology is described by Sarbin (1986) as the "storied nature of human conduct". Rather than existing as a theoretical perspective in its own right, narrative psychology embraces several theories, a common feature of which lies in a shared interest in identifying how human beings address their experiences through the construction of stories; and listening to the stories of others. Murray (1995) summarises the origin of narrative psychology as "a solution to the

problem of what stands between internal and external processes". Narrative measures could therefore be described as tools which contribute to the unravelling of the child's story within this context.

White (1990; 1995) is one of the founders of narrative therapy, and draws on a range of theoretical approaches in understanding the nature of language, conversation, and the way people make sense of their 'being in the world'. He describes the 'deconstruction' of narratives in therapy as central to the process of change. With children, this can take the form of play, using techniques which will enhance children's ability to describe and act out existing stories as well as defining alternative constructions of specific situations or interactions which are problematic e.g. through the use of puppets, De Amorium & Cavalcante (1993).

Clinical impressions of the TBP suggest it may provide a similarly useful therapeutic tool for use with young children within a narrative therapy framework. The method was successful in eliciting stories from children in the context of assessment in this and previously reported studies. In therapy, the stories could easily be transferred into individual working with children to enable them to tell their stories, and to reframe their solutions to hypothetical social and moral dilemmas to develop problem solving skills and enhance coping strategies. Developing individual stories during therapy, in addition to the 9 stems in the original measure would provide an opportunity to adapt the technique in accordance with clinical formulation. With further psychometric development, it could also be useful for assessing therapy outcome if administered pre and post intervention; as discussed in section 4.9 below.

4.9 What next? Further research using the TBP

In order to increase the discriminant validity of the TBP, several steps need to be taken to first of all improve the reliability of the measure. Work has already been undertaken by Sharp (1999) in a follow on study in the same local community with a sample of 59 children. The sample covered a wider age range (4-7 years), and the study was designed to assess concurrent and predictive validity of the measure, taking into consideration age and verbal ability.

Findings from this study were used to inform the methodology for the second study. Administration of the TBP was completed by Sharp throughout, which reduced tester and coder variability. Preliminary findings suggest that reliability is significantly

improved. Internal consistency of total positive and negative theme scores revealed alpha coefficients of .69 and .74. Factor analysis using 11 items revealed a 3 factor structure, which closely replicates that reported by Mueller (1996). Internal consistency of each factor based subscale score was acceptable for 2 factors (alpha coefficients = .71; .68; and .54). Results support the continued exploration of the measure.

Further work is planned in refining the coding procedure for the TBP. Data from both studies will be recoded (n=94) to include every occurrence of each theme in each story, with additional codes for use with mutually coded items (valuing/devaluing codes). Item scores which were not included in total or factor based subscale scores in the present and previously reported TBP studies will be included (distracted/impulsive, immature story), based on clinical findings in the present sample, and results from the study using the MUG (Mueller, in press).

Data collection from children referred to two clinical services is also being planned. Parents of children aged 4-6 years who are identified as 'at risk' or 'cases' by parents or teachers in the local whole population child mental health 'screen' will be approached to seek permission to administer the measure with their child. Findings from children identified by both informants, and those identified by one informant but not the other will be compared to establish similarities and differences in themes expressed, and total scores to see if the TBP contributes to the identification procedure.

Discussions are also underway in planning to administer the TBP as a pre and post measure with children who are currently receiving a 'universal' whole class preventive intervention in four local primary schools (Appleton & Hammond-Rowley, submitted paper). The intervention, Promoting Alternative THinking Strategies (PATHS: Greenberg & Kusche, 1994; 1997) is designed to improve children's emotion regulation and social competence. It is essentially a psycho-educational approach which has demonstrable positive effects at post intervention and follow up in the US (Greenberg, Kusche, Cook & Quamma, 1995). Emotional awareness and understanding are implicit themes throughout the intervention, which has four broad teaching objectives. These are to increase children's abilities to control arousal and behaviour through self regulation, to focus on affective vocabulary and emotion processing, to integrate affective, cognitive, and linguistic skills for effective problem

solving; and to promote positive self esteem and peer relationships.

Themes identified in the the TBP are directly relevant to the content of the PATHS Curriculum. The 9 stories which form the TBP will be retained in order to increase sampling of children using the original measure, but new stories will be constructed to tap specific themes and emotions which relate to material taught during the intervention. This will contribute to increasing understanding of the applicability of a 'mix and match' approach with the TBP, and also contribute to establishing outcomes of the whole class intervention.

Data collection is also being planned from children aged 4-6 years who are referred to a Child and Adolescent Mental Health Service (CAMHS). When clinical decisions are made to administer the CBCL and TRF (Achenbach, 1991) with adult informants, permission will be sought to collect concurrent TBP measures from children by a member of clinical staff to include children from a clinical sample. Two additional Child Therapists are currently being trained in the administration of the measure.

4.10 Conclusion

The associations between TBP total positive and negative scores, and concurrent parent and teacher measures were all in the expected direction; although statistically non significant. Results from the reliability analysis strongly suggest that the internal consistency of the total scores used to summarise children's responses on the TBP were unreliable in this study, which raises questions about the appropriateness of using summed total scores at all. It is difficult to draw conclusions from composite total scores when each is measuring more than one theme. Factor based subscale scores were equally unstable in this study.

Several methodological issues are likely to have contributed to the low internal consistency of the measure. These included specific difficulties which were encountered during the administration of the TBP. The sample size was smaller than expected, and specific rules for the administration of the TBP were difficult to adhere to consistently; despite training and practice during the pilot stage. Of particular relevance were the scoring directions. Several scoring rules were difficult to apply to the thematic material. The range of scores across the sample was limited, due to scoring the first only occurrence of any theme during each story, and the limited opportunity to score mutually coded themes (e.g. valuing/devaluing codes,

statements of relationship). Coding emotional expression was limited due to the directions associated with tester prompting. A large amount of information generated by the children during their stories was therefore not included in the total summary scores.

Refinement of the coding procedure to include more detailed criteria for scoring the presence of each theme, and the inclusion of each occurrence of each theme would increase the reliability and consistency of the coding process, and increase the range of scores generated by children from the measure. It may also be useful to include Likert scales to quantify the expression of the summary items, as used in the scoring of the MSSB.

Clinical impressions in this study, and further work by Mueller (in press) with older children strongly suggest that single items which are not included in total scores in the TBP are important in identifying children with difficulties. Through establishing reliability, further work will be required in establishing the validity of the TBP.

Comparison of findings with parent and teacher rated measures of child emotion and behaviour, and additional direct measures with children (e.g. peer nomination, observation) will allow conclusions to be drawn about the meaning of single item and composite total scores. It may then be possible to identify valid cut-off scores for the measure.

A considerable body of empirical evidence suggests that access to young children's internal representations of self and other is available through narrative story based techniques. This suggests that the child's own voice has the potential to be heard in the identification of emotional and behavioural problems, and in evaluating the effectiveness of intervention.

References

Aber, J. L., & Cicchetti, D. (1984). Socioemotional development in maltreated children: An empirical analysis. In H. Fitzgerald, B. Lester, & M. Yogman (Eds.), Theory and research in behavioural pediatrics (pp 147-205). New York: Plenum.

Achenbach, T.M. (1991). Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles. Burlington, VT: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1995). Epidemiological applications of multi-axial empirically based assessment and taxonomy. In F. C. Verhulst & H. M. Koot (Eds.), The epidemiology of child and adolescent psychopathology (pp.22-41). Oxford: Oxford University Press.

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical attempts. Psychological Bulletin **85** 1275-1301.

Achenbach, T. M., & Edelbrock, C. S. (1987). Manual for the youth self report and profile. Burlington: University of Vermont, Department of Psychiatry.

Ainsworth, M.B.S., Blehar, M., Waler E. & Wall, S. (1978). Patterns of attachment. Hillsdale NJ: Erlbaum.

Allen, J. P., Moore, C., Kuperminc, G., & Bell, K. (1998). Attachment and adolescent psychosocial functioning. Child Development, **69**, 1406-1419.

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E., & Rutter, M. (1995). The child and adolescent psychiatric assessment (CAPA). Psychological Medicine, **25**, 739-753.

Appleton, P. L., & Hammond-Rowley, S. (1999). Addressing the population burden of child and adolescent mental health problems: A primary care model. Submitted to Child Psychology and Psychiatry Review.

Appleton, P.L., Whitaker, C., Wilkinson, C., Hibbs, R., & Edwards, R.T. (1997). A population-based intervention for psychological problems in young children: A controlled trial and economic appraisal. Unpublished Grant Proposals to Wales Office of Research & Development for Health & Social Care.

Bee, H.L. (1992). The Developing Child (6th Ed). New York: Harper Collins.

Bellack, L. & Bellack S.S. (1974) Manual of instructions for the children's Apperception Test. New York: CPS Co.

Belsky, J., & Most, R. K. (1982). Infant exploration and play. In J. Belsky (Ed.), In the beginning New York: Columbia University Press.

Berger, M. (1994). Psychological tests and assessment. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and adolescent psychiatry. Modern approaches Oxford: Blackwell Science.

Bird, H.R., Gould, M.S., & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. J. AM Acad. Child Adolesc. Psychiatry, 31 78-85

Bolton, P., & Holland, A. (1994). Chromosomal abnormalities. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry. Modern Approaches (pp. 1 52-1 72). Oxford, London: Blackwell Science.

Bowlby, J. (1973). Attachment and loss. New York: Basic.

Bowlby, J. (1988). Developmental psychiatry comes of age. American Journal of Psychiatry, 45, 1-10.

Boyle, M. H., & Jones, S. C. (1985). Selecting measures of emotional and behavioural disorders of childhood for use in general populations. Journal of Child Psychology & Psychiatry, 26, 137-159.

Bretherton, I. (1990a). Open communication and internal working models: Their role in attachment relationships. In R. Thompson (Ed.), Nebraska Symposium on Motivation, 36 (pp. 59-113). Lincoln, Nebraska: University of Nebraska Press.

Bretherton, I. (1990b). The MacArthur story-stem battery (Unpublished manual) The MacArthur Research Network for Early Childhood Transitions.

Bretherton, I., and Bates, E. (1985). The development of representation from 10 to 28 months: Differential stability of language and symbolic play. In R. Emde & R. Harmon. (Eds.), Continuities and discontinuities in development (pp. 171-199). New York: Plenum

Bretherton, I., Ridgeway, D., & Cassidy, J. (1990). The role of internal working models in the attachment relationship. In M. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), Attachment during the preschool years (pp. 3-39). Chicago: Chicago University Press.

Bruner, J. (1986). Actual minds. possible worlds. Cambridge, MA: Harvard University Press.

Buchsbaum, H. K., & Emde, R. N. (1990). Play narratives in 36 month old children: Early moral development and family relationships. Psychoanalytic study of the child, 45, 129-155.

Buchsbaum, H. K., Toth, S. L., Clyman, R. B., Cicchetti, D., & Emde, R. N. (1992). The use of narrative story stem technique with maltreated children. Development & Psychopathology, 4, 603-625.

Caplan, G. (1964). Principles of preventative psychiatry. New York: Basic Books.

Carlson, V., Cicchetti, D., Barnett, D., & Braunwald, K. G. (1989). Disorganized/ disoriented attachment relationships in maltreated infants. Developmental Psychology, 25(4), 525.

Chazan, M. (1970). Inconsequential behaviour in school children. Brit. J. Educ. Psychol. 40, 5-7.

Cicchetti, D., & Barnett, D. (1991). Attachment organisation in preschool aged maltreated children. Development & Psychopathology, 3, 397-411.

Cicchetti, D., Ganiban, J., & Barnett, D. (1991). Contributions from the study of high risk populations to understanding the development of emotion regulation. In J. Garber & K. A. Dodge (Eds.), The development of emotion regulation and dysregulation (pp. 15-48). New York: Cambridge University Press.

Clements, J. (1998). Training parents of mentally handicapped children. In N. Spencer (Ed). Progress in community child health Vol. ii Edinburgh: Churchill Livingstone.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. Science(221), 207-210.

Cohen, P., Velez, N., Kohn, M., Schwab-Stone, M., & Johnson, J. (1987). Child psychiatric diagnosis by computer algorithms: Theoretical issues and empirical tests. Journal of the American Academy of Child & Adolescent Psychiatry, 26, 631-638.

Coie, J. D., & Dodge, K. A. (1988). Multiple sources of data on social behaviour and social status in the school: a cross age comparison. Child Development, 59, 815-829.

Cole, P.M., Michel, M.K., & O'Donnell-Teti, L. (1994). The development of emotion regulation and dysregulation: A clinical perspective. Monographs of the Society for Research in Child Development, 59 (2-3, 240).

Collins, N.L. & Read, S.J. (1990). Adult attachment, working models and relationship quality in dating couples. Journal of Personality and Social Psychology 58, 4, 644-663.

Comrey, A., & Lee, H.B. (1992). A first course in factor analysis. (2nd Ed.) Hillsdale, N.J : Erlbaum.

Crittenden, P. M. (1988). Relationships at risk. In J. Belsky & T. Nezworski (Eds.), Clinical implications of attachment theory (pp 136-174) Hillsdale, NJ: Erlbaum.

De Amorium, A.C. & Cavalcante, F.G. (1993). Narrations of self: Video production in a marginalized subculture. In S. MacNamee, & J. Gergen (Eds.), Therapy as Social Construction. London, Sage.

Dowdney, L., Mrazek, D., Quinton, D., & Rutter, M. (1984). Observation of parent-children interaction with two to three year olds. Journal of Child Psychology & Psychiatry, 25, 379-407.

Dunn, J., & Dale, N. (1984). I a daddy: 2 year old's collaboration in joint pretend with sibling and mother. In I. Bretherton (Ed.), Symbolic play. The development of social understanding (pp. 131-158). Orlando: Academic Press Inc.

Dunn, J., Brown, J. R., & Maguire, M. (1995). The development of children's moral sensibility: Individual differences and emotion understanding. Developmental Psychology, 31, 649-659.

Dunn, L. M., & Dunn, L. (1981). The Peabody Picture Vocabulary Test -Revised. Circle Pines: MN: American Guidance Service.

Edelbrock, C., Costello, A., Dulcan, M.K., Conover, N.C., & Kalas, R. (1986). Parent - child agreement on child psychiatric symptoms assessed via structured interview. Journal of Child Psychology & Psychiatry, 27, 181-190

Eisenberg, N., & Miller, P.A. (1987). The relation of empathy to prosocial and related behaviors. Psychology Bulletin, 101, 91-119.

Elander, J., & Rutter, M. (1995). Use and development of the Rutter Parent's and teacher's scales. International Journal of Methods in Psychiatric Research, 5, 1-16.

Emde, R., Birgen, Z., Clyman, R., & Oppenheim, D. (1991). The moral self of infancy: Affective core and procedural knowledge. Developmental Review, 11, 251-270.

Emde, R., Kubicek, L., Oppenheim, C., & Oppenheim, O. (1997). Imaginative reality observed during early language development. International Journal of Psychoanalysis 78, 115-133.

Emery, R. E. (1982). Interparental conflict and the children of discord and divorce. Psychological Bulletin, 92, 310-330.

Farver, J. (1992). Communication and shared meaning in social pretend play. Early Childhood Research Q, 7, 501-516.

Field, T. (1985) Attachment as psychobiological attunement: Being on the same wavelength. In M. Reife & T. Field (Eds.), The psychobiology of attachment & separation (pp 415-454). Orlando: Academic Press

Fisher, E.B. (1995). Editorial: The results of the COMMIT trial. American Journal of Public Health, 85, 159-160.

Fivush, R. (1991). The social construction of personal narratives. Merrill-Palmer Quarterly, 37, 59-82.

Flavell, J. H. (1986). Development of knowledge about the appearance-reality distinction. Monographs of the Society for Research in Child Development, (Serial No.212).

Floyd, F.J. & Zmich, D.E. (1991). Marriage and the parenting partnership: Perceptions and interactions of parents with mentally retarded and typically developing children. Child Development, 62, 1434 - 1448

Gaensbauer, T. J., & Harmon, R. J. (1982). Attachment behavior in abused/neglected and premature infants: Implications for the concept of attachment. In R. N. Emde & R. J. Harmon (Eds.), Attachment and affiliative systems (pp. 245-279). New York: Plenum.

Gardner, F. E. M. (1994). The quality of joint activity between mothers and their children with behaviour problems. Journal of Child Psychology & Psychiatry, 35, 1483-1494.

- Gardner, M. D. (1990). Expressive One-Word Picture Vocabulary Test (Revised). Novato, CA: Academic Therapy Publications.
- Garfield, S. L. (1974). Clinical psychology. New York: Pergamon Press.
- Garnezy, N., & Masten, A. S. (1994). Chronic adversities. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry. Modern Approaches (pp. 191-209). Oxford, London: Blackwell Science.
- Glasberg, R., & Aboud, F. (1982). Keeping one's distance from sadness: Children's self reports of emotional experience. Developmental Psychology 18, 287-293.
- Gleser, G. (1963). Projective methodologies. Ann. Rev. Psychol., 14. Palo Alto, California: Annual Reviews Inc.
- Golumb, C., & Balasso, L. (1995). Make believe and reality: Explorations of the imaginary realm. Developmental Psychology, 31, 800-810.
- Goodman, R. (1994a). Brain disorders. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry. Modern Approaches (pp. 172-191). Oxford, London: Blackwell Science.
- Goodman, R. (1994b). A modified version of the Rutter parent questionnaire including items on children's strengths: A research note. Journal of Child Psychology & Psychiatry, 35, 1483-1494.
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. Journal of Child Psychology & Psychiatry, 38, 581-586.
- Goyette, C. H., Conners, K. C., & Ulrich, F. (1978). Normative data on revised Conners parent and teacher rating scales. Journal of Abnormal Child Psychology, 221-236.
- Greenberg, Kusche, Cook, Quamma (1995) Promoting emotional competence in school-aged children: The effects of the PATHS curriculum. Development & Psychopathology, 7, 117-136

Greenberg, M.T., Kusche, C.A. (1997). Improving children's emotion regulation and social competence: The effects of the PATHS curriculum. Paper presented at the Society for Research in Child Development, Washington DC., April.

Griffiths, R. (1970). The abilities of young children. London University of London Press.

Hammond-Rowley, S., & Appleton, P. L. (1997). A community based child mental health needs assessment: The key to effective prevention? Paper presented at 'Options for Life: Seventh Annual European Conference on the Promotion of Mental Health, . Maastricht: 8 - 10th October.

Harris, P., & Kavanaugh, R. (1993). Young children's understanding of pretense. Monographs of the Society for Research in Child Development, 58 (Serial No. 231).

Herjanic, B., & Reich, W. (1982). Development of a structured psychiatric interview for children: Agreement between child and parent on individual symptoms. Journal of Abnormal Child Psychology, 25(1), 21-31.

Hersov, L. (1994). Adoption. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry. Modern Approaches (pp.267-285). Oxford, London: Blackwell Science.

Hinshaw, S. P., Han, S. S., Erhardt, D., & Huber, A. (1992). Internalizing and externalizing behaviour problems in preschool children: Correspondence among parent and teacher rating and behaviour observations. Journal of Clinical Child Psychology, 21, 143-150.

Hoffman, M.L. (1983). Affective and cognitive processes in moral internalization. In E.T. Higgins, D.N. Ruble & W.W. Hartup (Eds). Social cognition and social development: A sociocultural perspective. (pp 236 - 274). New York: Cambridge University Press.

Home Office (1991). Working together under the Children Act 1989 Department of Health, Department of Education and Science, and Welsh Office.

Hudson, J. A. (1990). The emergence of autobiographic memory in mother-child conversation. In R. Fivush & J. A. Hudson (Eds.), Knowing and remembering in young children (pp. 166-196). Cambridge, England: Cambridge University Press.

Institute of Medicine (1994). Reducing risks for mental disorders. Frontiers for preventive intervention research. Washington DC: National Academy Press.

Jenkins, J.M., & Smith, M.A. (1990). Factors protecting children in disharmonious homes: Maternal reports. Journal of the American Academy of Child & Adolescent Psychiatry, 29, 60-69.

Kaufman, J., & Cicchetti, D. (1989). The effects of maltreatment on school aged children's socioemotional development: Assessments in a day camp setting. Developmental Psychology, 25, 516-524.

Kelly, G. A. (1955). Psychology of personal constructs. New York: Norton.

Kochanska, G., & Aksan, N. (1995). Mother-child mutually positive affect, the quality of child compliance to request and prohibitions, and maternal control as correlates of early internalization. Child Development, 66, 236-254

Kochanska, G., DeVet, K., Goldman, M., Murray, K., & Putnam, S. P. (1994). Maternal reports of conscience development and temperament in young children. Child Development, 65, 852-868.

Kochanska, G., Padavich, D. L., & Koenig, A. L. (1996). Children's narratives about hypothetical moral dilemmas and objective measures of their conscience: Mutual relations and socialization antecedents. Child Development, 67, 1420-1436.

Kolko, D.J., Kazdin, A.E. (1993). Emotional/behavioural problems in clinic & non clinic children: Corresspondence among child, parent and teacher reports. Journal of Child Psychology & Psychiatry 34, 991-1006.

Kolvin, I., Charles, G., Nicholson, K., Fleeting, M., Fundundis, T. (1990). Factors in prevention in inner-city deprivation. In D. Goldberg & D. Tantom (Eds) The Public Health Impact of Mental Disorder, pp 115-123, Toronto: Hogrefe & Huber.

Kolvin, I., Miller, F., Fleeting, M. & Kolvin, P.A. (1988). Social and parenting factors affecting criminal - offence rates. Findings from the Newcastle Thousand Family study (1947-1980). British Journal of Psychiatry, 152, 80-90.

Kovacs, M. (1982) The Children's Depression Inventory. Unpublished manuscript, University of Pittsburgh.

La Greca, A.M., & Silverman, W.K. (1993). Parent reports of child behaviour problems: Bias in participation. Journal of Abnormal Child Psychology, 21, 89-101

Lewis, M., & Brooks-Gunn, J. (1979). Social cognition and the acquisition of self. New York: Plenum.

Maccoby, E. (1980). Social development. New York: Harcourt, Brace, Jovanovich.

Main, M., & Hesse, P. (1990). Parent's unresolved traumatic experiences are related to infant disorganized attachment status: Is frightened and/or frightening parent behavior the linking mechanism? In M. Greenberg, D. Cicchetti, & E. M. Cummings (Eds.), Attachment during the preschool years (pp. 161-182). Chicago: Chicago University Press.

Mayes, L. C., & Cohen, D. J. (1993). Playing and therapeutic action in child analysis , Journal of Psychoanalysis 74, 1235-1244.

McCabe, A., & Peterson, C. (1991). Getting the story: A longitudinal study of parental styles in eliciting narratives and developing narrative skill. In A. McCabe & C. Peterson (Eds.), Developing narrative structure Hillsdale, NJ: Erlbaum.

McCune, L. (1995). A normative study of representational play at the transition to language. , 31, 198-206.

McGuire, J., & Richman, N. (1986). Pre-school behaviour checklist handbook. Windsor, Berkshire: NFER Nelson.

Ministry of Education (1955). Report of the Committee on maladjusted children. London: HMSO.

Mueller, E., & Tingley, E. (1990). The bear's picnic: Children's representations of themselves and their families. In I. Bretherton & M.W. Watson (Eds.), Children's perspectives on the family. San Francisco: Jossey-Bass.

Mueller, N. (in press) The MUG and the Teddy Bear's picnic: Story-telling techniques for the assessment of children's emotional and behavioural problems. In K. Gitlin-Weiner, A. Sandgrund, & C. E. Schaeffer (Eds.), Play diagnosis and assessment. Revised Edition New York: John Wiley

Mueller, N. (1996). The teddy bear's picnic: Four year old children's personal constructs in relation to behavioural problems and to teacher global concern. Journal of Child Psychology & Psychiatry, 37(4), 381-389.

Murray, H.A. (1943). Thematic Apperception Test. Cambridge MA: Harvard University Press.

Murray, K.D. (1995) Narratology. In J.A. Smith, R. Harre, & L. Van Langenhove (Eds.), Rethinking Psychology London, Sage

NHS/Health Advisory Service (1995). Child and adolescent mental health services: Together we stand. The commissioning role and management of child and adolescent mental health services. HMSO.

Offord, D.R., Boyle, M.H., Racine, Y., Szatmari, P., Fleming, J.E., Sanford, M., & Lipman, E.L. (1996). Integrating assessment data from multiple informants, J. Am. Acad. Child Adolesc. Psychiatry, 35(8), 1078-1085.

Oppenheim, D., Nir, A., Warren, S., & Emde, R. N. (1997). Emotion regulation in mother-child narrative co-construction: Associations with children's narratives and adaptation. Developmental Psychology, 33(2), 284-294

Patrick, M. (1994). Personality disorder and the mental representations of early social experience. Development & Psychopathology, 6, 375-388.

Pavuluri, M.N., Luk, S.L., & McGee, R. (1996). Help seeking for behaviour problems by parents of preschool children: A community study. Journal of the American Academy of Child & Adolescent Psychiatry, 35, 215-222.

Perrin, S., & Last, C. G. (1992). Do childhood anxiety measures measure anxiety? Journal of Abnormal Child Psychology, 20, 567-578.

Raven, J.C. (1951). Controlled projection for children (2nd Ed.). London: H.K. Lewis.

Reese, E., & Fivush, R. (1993). Parental styles of talking about the past. Developmental Psychology, 29 596-606.

Richman, N., Stevenson, J., & Graham, P. (1982). Pre-school to school. A behavioural study. London: Academic Press.

Rothbart, M. K., Ahadi, S. A., & Hershey, K. L. (1994). Temperament and social behavior in childhood. Merrill-Palmer Quarterly, 40, 21-39.

Rotter, J.B. & Rafferty, J.E. (1950). Manual for the Rotter incomplete sentences blank, College Form. New York: Psychological Corporation

Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers. Journal of Child Psychology & Psychiatry, 8, 1-11.

Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistance to psychiatric disorder. British Journal of Psychiatry, 147, 598-611.

Rutter, M. (1989). Isle of Wight revisited: Twenty five years of child psychiatric epidemiology. J. Am Acad. Child Adolesc Psychiatry, 28, 633-653.

Rutter, M. (1990). Competence under adversity: Individual and family differences in resilience. In J. Rolf, A. S. Masten, D. Cicchetti, K. H. Nuechterlein, & S. Weintraub (Eds.), Risk and protective factors in the development of psychopathology New York: Cambridge University Press.

Rutter, M. (1993). Cause and course of psychopathology: Some lessons from longitudinal data. Paediatric & Perinatal Epidemiology 7, 105-120.

Rutter, M., & Smith, D. J. (1995). Psychosocial disorders in young people. Time trends and their causes. Chichester: John Wiley & Sons.

Rutter, M., Tizard, J., & Whitmore, K. (1970). Education, health and behaviour. London: Longman.

Sarbin, T.R. (1986) (Ed.). Narrative Psychology: The storied nature of human conduct. New York: Praeger.

Schachar, R., Rutter, M., & Smith, A. (1981). The characteristics of situationally and pervasively hyperactive children: Implications for syndrome definition. Journal of Child Psychology & Psychiatry, 22, 375-392.

Schaffer, H. R. (1990). Making decisions about children: Psychological questions and answers. Oxford, UK: Blackwell.

Schneidman, E.S. (1947). Make a Picture Story (MAPS). New York: Psychological Corporation

Schore, A. N. (1994). Affect regulation and the origin of the self: The neurobiology of emotional development. Hillsdale, NJ: Erlbaum.

Scott, S. (1996). Measuring oppositional and aggressive behaviour. Child Psychology & Psychiatry Review, 1(3), 104-109.

Sharp, H. (1999). TBP: Preliminary findings from a study with 5-8 year olds. Personal communication

Shepherd, M., Oppenheim, A.N., & Mitchell, S. (1991). Childhood behaviour and mental health. London: University of London Press.

Simonoff, E., McGuffin, P., & Gottesman, I. I. (1994). Genetic influences on normal and abnormal development. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry. Modern Approaches (pp. 129-152). Oxford, London: Blackwell Science.

Singer, D. G., & Singer, J. L. (1990). The house of make-believe. Cambridge, MA: Harvard University Press.

Skuse, D., & Bentovim, A. (1994). Physical and emotional maltreatment. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry. Modern Approaches (pp. 209-230). Oxford, London: Blackwell Science.

Smith, J.A. (1995) Qualitative methods, identity and transition to motherhood. The Psychologist, 8, 122-125.

Smith, M., & Bentovim, A. (1994). Sexual abuse. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry. Modern Approaches (pp. 230-252). Oxford, London: Blackwell Science.

Smith, P. K. (1990). Rough and tumble play, aggression and dominance: Perception and behavior in children's encounters. Human Development(33), 271-282.

Sorce, J. F., & Emde, R. N. (1981). Mother presence is not enough. Developmental Psychology, 11(6), 737-745.

Sroufe, L. A. (1989). Pathways to adaptation and maladaptation: Psychopathology as developmental deviation. In D. Cicchetti (Ed.), Rochester Symposia on Developmental Psychopathology (pp. 13-14). Hillsdale, NJ: Enbaum.

Sroufe, L. A. (1991). Considering normal and abnormal together: The essence of developmental psychopathology. Development and psychopathology(2), 335-347

Sroufe, L. A. (1996). Emotional development. The organization of emotional life in the early years. New York: Cambridge University Press.

Sroufe, L. A., Carlson, E., & Shulman, S. (1993). Individuals in relationships: Development from infancy through adolescence. In D. C. Funder, R. D. Parke, C. Tomlinson-Keasey, & K. Widaman (Eds.), Studying lives through time: Personality and Development (pp. 315-342). Washington DC: APA.

Stacey, M., Dearden, R., Pill, R., Robinson, D. (1970). Hospitals, children and their families: The report of a pilot study. London: Routledge.

Stallard, P. (1995). New thoughts on the management of behaviour problems in pre-school children. In N. Spencer (Ed.), Progress in community child health (pp. 93-107). Edinburgh: Churchill Livingstone.

Stanger, C. & Lewis, M. (1993). Agreement among parents, teachers and children on internalizing and externalising behavior problems. Journal of Clinical Child Psychology, 22, 107-115

Stevens, J. (1992). Applied multivariate statistics for the social sciences (2nd Ed), Hilldale, N.J.: Erlbaum.

Sturgess, J., & Ziviani, J. (1995). Development of a self-report play questionnaire for children aged 5 to 7 years: A preliminary report. Australian Occupational Therapy Journal, 42, 107-117.

Susser, M. (1995). Editorial: The tribulations of trials - intervention in communities. American Journal of Public Health, 85, 157-158.

Tabachnick, B.G., & Fidell, L.S. (1996). Using multivariate statistics (3rd Ed.). New York: Harper Collins.

Tamis-LeMonda, S. (1992). Language and play at one year: a comparison of toddlers and mothers in the United States and Japan. International Journal of Behaviour Development, 15, 19-42.

Taylor, E., Schachar, R., Thorley, G., & Weisberg, M. (1986). Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. British Journal of Psychiatry, 149, 760-777

Tharinger, D. J., & Stark, K. (1990). A qualitative vs. quantitative approach to evaluating the draw-a-person and kinetic family drawing: A study of mood-and anxiety-disorder children. Psychological Assessment, 4 365-375.

Troy, M., & Sroufe, L. A. (1987). Victimization among preschoolers: The role of attachment relationships. J. Am. Acad. Child Adolesc. Psychiatry, 26, 166-172.

Urban, J., Carlson, E., Egeland, B., & Sroufe, L. A. (1991). Patterns of individual adaptation across childhood. Development and Psychopathology(3), 445-460.

Verhulst, F. C., & Ende, J. V. D. (1992). Agreement between parents reports and adolescents self reports of problem behaviour. Journal of Child Psychology & Psychiatry, 33(6), 1011-1023.

Verhulst, F. C., & Koot, H. M. (1995). The epidemiology of child and adolescent psychiatry. Oxford: Oxford University Press.

Warren, S. L., Oppenheim, D., & Emde, R. N. (1996). Can emotions and themes in children's play predict behavior problems? J. Am. Acad. Child Adolesc. Psychiatry 34(10), 1331-1337.

Warren, S., Emde, R. N., & Oppenheim, D. (1993). New MacArthur emotion story stems . MacArthur Research Network for Early Childhood Transitions.

Warren, S., Mantz-Simmons, L., & Emde, R. N. (1993). Narrative emotion coding manual MacArthur Foundation Network on Early Childhood Transitions.

Wechsler, D. (1967). Manual for the Wechsler Preschool and Primary Scale of Intelligence. San Antonio, TX: Psychological Corporation.

Wechsler, D. (1974). Manual for the Wechsler Intelligence Scale for Children - Revised. New York: Psychological Corporation.

Weinberger, L. A., & Starkey, P. (1994). Pretend play by African American children in Head Start. Early Childhood Research Q(9), 327-343.

Weisz, J.R., Suwanlert, S., Chaiyasit, W., & Walter, B.R. (1987). Over and under controlled referral problems among children and adolescents from Thailand and the United States: The wat and wai of cultural differences. Journal of Consulting & Clinical Psychology, 55, 719 - 726.

Welsh Office (1997). The BEST for Special Education: A Welsh Office Green Paper on Special Educational Needs. Cardiff: Welsh Office.

Werner, E. E., & Smith, R. S. (1992). Overcoming the odds. High risk children from birth to adulthood. New York: Cornell University Press.

White, M. & Epsom, D. (1990). Narrative means to therapeutic ends. New York: Norton.

White, M. (1995). Re-authoring lives: Interviews & Essays Adelaide: Dulwich Centre Publications.

Wilson, B., Ivani-Chalian, R., & Aldrich, F. (1991). The Rivermead Behavioural Memory Test for Children aged 5 - 10 years. Bury St Edmonds: Thames Valley Test Company.

Wolf, D. P (1984). Agency and experience: Actions and states in play narratives. In I. Bretherton (Ed.), Symbolic play. The development of social understanding (pp. 195-217). Orlando: Academic Press Inc.

Wolf, D. P., & Slade, A. (1994). Modes of meaning: Clinical and developmental approaches to children's play. Oxford: Oxford University Press.

Wolkind, S., & Rushton, A. (1994). Residential and foster family care. In M. Rutter, E. Taylor, & L. Hersov (Eds.), Child and Adolescent Psychiatry Modern Approaches (pp. 252-267). Oxford, London: Blackwell Science.

World Health Organisation (1972). World Health Statistics Annual. Geneva.

World Health Organisation (1982) Changing patterns in suicide behaviour. Copenhagen, Denmark: WHO/Euro Reports & Studies.

Zazlow, M.J., & Hayes, C.D. (1986). Sex differences in children's response to psychosocial stress: Towards a cross context analysis. In M. Lamb, A. Brown, & B. Rogoff (Eds.), Advances in developmental psychology, Vol 4. Hilldale NJ: Erlbaum.

PAGE/PAGES
EXCLUDED
UNDER
INSTRUCTION
FROM
UNIVERSITY