"Small Talk"

Young children in the general population as informants of their problem behavior

Ank Ringoot

“Small Talk”:

Young children in the general population as informants of their problem behavior

Ank Ringoot

"Small Talk": Young children in the general population as informants of their problem behavior.
PhD thesis, Erasmus University Rotterdam, The Netherlands

Acknowledgements: The Generation R Study is conducted by Erasmus Medical Center Rotterdam in close collaboration with the Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam, and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR), Rotterdam. We gratefully acknowledge the contribution of children and parents, general practitioners, hospitals, midwifes and pharmacies in Rotterdam. The general design of the Generation R is made possible by the Erasmus Medical Center Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Netherlands Organization for Scientific Research (NWO), the Ministry of Health, Welfare, and Sport, and the Ministry of Youth and Families. The work presented in this thesis was conducted at the Department of Child and Adolescent Psychiatry/Psychology in close collaboration with Riagg Rijnmond. The work at the Department of Child and Adolescent Psychiatry/Psychology was supported by a grant from ZonMW (Grant Number: 10.000.1003). Financial support for the publication of this thesis was provided by the Department of Child and Adolescent Psychiatry/Psychology, the Generation R Study, and the Erasmus University Rotterdam.

ISBN: 978-94-6259-635-1
Cover photo: Alain Gil-Gonzalez via Generation R
Cover design and layout: Hans Ringoot
Printed by: Ipskamp drukkers, Enschede, The Netherlands

"Small Talk": Young children in the general population as informants of their problem behavior

Kinderen aan het woord: Jonge kinderen uit de algemene populatie als informant over hun probleemgedrag

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 3 juni 2015 om 11.30 uur

door

Ank Paulien Ringoot
geboren te Hulst

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof.dr. H. Tiemeier
	Prof.dr. F.C. Verhulst
Overige leden:	Prof.dr. J.J. van Busschbach
	Prof.dr. L. Arseneault
	Prof.dr. M.H. van IJzendoorn
Copromotor:	Dr. P.W. Jansen

Paranimfen:	Nina van Mil
	Renske van Wijk

CONTENTS

MANUSCRIPTS BASED ON THIS THESIS

Chapter 2:

Ringoot, A.P., Jansen, P.W., Steenweg-de Graaff, J., Measelle, J.R., van der Ende, J., Raat,H., Jaddoe, V.W.V., Hofman, A., Verhulst, F.C., & Tiemeier, H. (2013) Young children's self-reported emotional, behavioral and peer problems: the Berkeley Puppet Interview. *Psychological Assessment, 25*(4), 1273-1285.

Chapter 3:

Ringoot, A.P., Jansen, P.W., Rijlaarsdam, J., So, P., Jaddoe, V.W.V., Hofman, A., Verhulst, F.C., & Tiemeier, H. Self-reported problem behavior in young children with and without a DSM-based diagnosis in the general population. *Submitted for publication*.

Chapter 4:

Ringoot, A.P., Ende, van der, J., Jansen, P.W., Measelle, J.R., Basten, M., So, P., Jaddoe, V.W.V., Hofman, A., Verhulst, F.C., & Tiemeier, H. (2015) Why mothers and young children agree or disagree in their reports of the child's problem behavior. *Child Psychiatry and Human Development.* Advance online publication.

Chapter 5:

Ringoot, A.P., Tiemeier, H., Jaddoe, V.W.V., So, P., Hofman, A., Verhulst, F.C., & Jansen, P.W. (accepted for publication). Parental depression and child well-being: young children's self-reports helped addressing biases in parent reports. *Journal of Clinical Epidemiology*.

Chapter 6:

Mackenbach, J.D., Ringoot, A.P., Ende, van der, J., Verhulst, F.C., Jaddoe, V.W.V., Hofman, A., Jansen, P.W., & Tiemeier, H. (2014) Exploring the relation of harsh parental discipline with child emotional and behavioral problems by using multiple informants: the Generation R Study. *PlosOne*, *9*(8), e104793.

Chapter 7:

Ringoot, A.P., Jansen, P.W., Kok, R., Verlinden, M., Jaddoe, V.W.V., Verhulst, F.C., & Tiemeier, H. Parenting, children's self-regulation and the quality of their peer relationships: the Generation R Study. *To be submitted.*

chapter 1

introduction

1

INTRODUCTION

In the Netherlands, no validated instruments are available to systematically assess young children's self-reported problem behavior. The studies presented in this thesis introduce and examine an interview that may enable researchers and clinicians to systematically obtain reports of emotional and behavioral problems from children aged four to eight years.

The curious case of psychological assessment

Psychologists and psychiatrists are faced with a difficult situation when assessing problem behavior. Unlike most medical disciplines, where pathology can be judged using relatively objective measures such as direct observations of weight and height, laboratory measurements of tissue, or imaging techniques, no objective or let alone gold standard measures are available to assess psychological problems. Instead, one has to rely largely on the assessed person's narrative. Hence, information is obtained by questionnaires, interviews and if possible by observations of behavior. Over the years, standardized assessment procedures have been developed (e.g. Achenbach et al., 1987a), that aid in interpreting problem behavior along uniform frameworks. Yet, the fact remains that characteristics vital to understand psychological problems are difficult to measure (Kraemer et al., 2003).

A particularly difficult situation pertains to the assessment of psychological problems in children. Rather than asking children about their feelings and behavior, it has long been common practice of child psychologists and psychiatrists to obtain information solely from parents (mainly mothers), teachers and observations. It was only until child reports were used in epidemiological studies, like the Isle of Wight study, that the value of children as informant with respect to their own well-being was acknowledged (Rutter, 1989; Rutter & Stevenson, 2008). Before that time, parents were assumed to be the best informants regarding their child's behavior and children's own information was considered inferior (Edelbrock et al., 1985). This view changed significantly in the last decades and today, it is widely accepted that it is useful to obtain children's views on problem behavior by standardized questionnaires and interviews (Grills & Ollendick, 2002). Accordingly, an increasing number of studies has been published using children as informants on a variety of outcomes (van de Looij-Jansen, 2010). Yet, the youngest children (younger than eight years of age) are still faced with the situation that their perspective on psychological problems is usually not taken into account in standardized assessments of child problem behavior.

This sole reliance on other informants than the young child seems odd, but it is also conceivable considering that obtaining information from young children is complicated by a variety of developmental factors (Measelle et al., 1998). For instance, before children are able to reveal their perception of emotional and behavioral problems, they should have developed a multidimensional view of themselves, known as a self-concept. A self-concept is a coherent, integrated knowledge structure that incorporates differentiated ideas someone has about him or herself (Baumeister, 2010). Children between four and eight were not thought to have differentiated self-concepts. Rather, the general belief was that they were only able to describe themselves in positive ways and in terms of concrete behavior (e.g. I am a sweet kid, I can run very fast) (Baumeister, 2010),

which would preclude obtaining reports about emotions and more complex behavior. However, studies in the past decades have shed new light on this issue, showing that children from four years onwards possess more differentiated concepts of themselves than was long assumed (Eccles et al., 1993; Goodvin et al., 2008; Marsh et al., 1991; Marsh et al., 2002). Young children can even describe their feelings, thoughts and intentions, as long as information is gathered with structured and age-appropriate methods (Ablow et al., 1999; Arseneault et al., 2005).

However, how to obtain information from young children in age-appropriate ways? To do so, several developmental factors that constrain the possibility to elicit reliable reports from young children should be overcome. These include young children's short attention span, their less mature expressive skills, potential anxiety to disclose information to strangers, and importantly, sensitivity to suggestion (Measelle et al., 1998). To increase task attention, children benefit from the use of toys like puppets, or other visual aids like pictures or cartoons (Eder, 1990; Valla et al., 2000). Further, methods should not rely exclusively on children's verbal abilities, as they are less likely to work with inhibited or language-delayed children (Measelle et al., 1998). In addition, young children tend to be susceptible to socially desirable responding or to suggestion (Ceci & Bruck, 1993). Literature has shown that as a result of suggestive interviewing children provide false claims of a range of both positive and negative topics (Bruck et al., 2002). This may be avoided by using standardized interviews and by using bipolar options (presenting both positive and negative poles in the same item) that are equally acceptable (Measelle et al., 1998). Finally, children appear to respond more honestly in situations where the syntactic structure and lexicon is adapted to the child's level (Eder, 1989).

Not one informant has a monopoly of wisdom

As a result of a the complicating factors described above, it seems no surprise that including young children's reports in standardized assessment of problem behavior is not yet widespread. Nevertheless, young children's perspectives are important to consider.

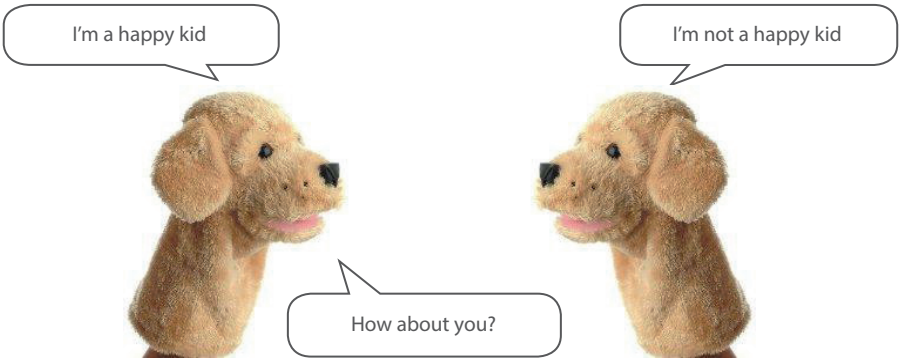
Assessments of child problem behavior based solely on adult informants may not be sufficient to obtain a complete description of children's problems, because every informant judges the same behavior differently and because the children's behavior may be context specific (De Los Reyes & Kazdin, 2005; Verhulst & van der Ende, 2008). Kraemer et al. (2003) therefore stated that a more complete picture of child functioning can only be obtained by using multiple informants. Importantly, informants should be carefully selected: they should not necessarily give the same information, rather they should provide information from different angles. Kraemer et al. (2003) compared this with a GPS system. One satellite will only give a rough estimation of your exact position, but once information from multiple satellites is included, the estimation of your position becomes more precise. If we translate this metaphor to assessment of child problem behavior, it means that instead of only considering information from the adult's perspective, information from other perspectives, like that of the child, may be vital for precise estimation of the problem. Indeed, if children suffer from problems like depression and anxiety, but do not show overt sadness, anhedonia or agitation, children's information is important as these problems are difficult

to recognize for adults (Achenbach et al., 1987b; Wu et al., 1999). In addition, children's behavior may vary depending on the context. A child can be disruptive at home, but able to inhibit him or herself at school, or the other way around. Although children too may have their blind spots, only children themselves have unlimited access to their inner world and can report about their behavior across different contexts. Thus, all informants seem to provide imperfect measures of child problem behavior, but each may also document valid aspects of children's behavior (Arseneault et al., 2005). To improve the accuracy of information concerning child problem behavior, not only the adult's, but also young children's reports should be considered.

The Berkeley Puppet Interview

One promising instrument that can be used to obtain information from young children is the Berkeley Puppet Interview (Ablow & Measelle, 2003), developed in the United States to address the absence of standardized measures to obtain young children's reports of problem behavior. The Berkeley Puppet Interview (BPI) is a short, structured, but interactive interview. The interview uses two hand puppets that make opposing statements about themselves to playfully discuss issues like emotional and behavioral problems, social functioning, and school functioning in children between four and eight years old (Ablow & Measelle, 2003; Ablow et al., 1999; Measelle et al., 1998). The interview is videotaped and later coded. The BPI was developed taking into account factors that might complicate obtaining reports of young children, such as their short attention span, less developed vocabulary, and stranger anxiety (Measelle et al., 1998). This was done by using puppets (Eder, 1990; Irwin, 1985), by including the possibility to respond non-verbally to the puppets' statements, and by the use of bipolar options to prevent suggestive interviewing (illustrated in Figure 1).

Figure 1. Illustration of the Berkeley Puppet Interview



Despite the development of instruments, like the BPI, to obtain reports about emotional and behavioral problems from young age onwards, the use of such interviews in research and clinical practice is rare. Moreover, in the Netherlands no validated instrument was available to obtain reports of emotional and behavioral problems from young children themselves. Consequently,

many questions concerning the use of young children's reports of emotional and behavioral problems are yet unanswered. For instance, although several studies using the BPI suggested that information obtained with the BPI is associated with external measures like observed classroom behavior, DSM-based classifications and future behavioral problems (Ablow et al., 2009; Arseneault et al., 2005), most of these studies were performed in small, homogeneous samples or on a limited set of problem domains. Further, as discussed, informants often differ in the ways they judge behavior, resulting in discrepancies (De Los Reyes & Kazdin, 2005; Verhulst & van der Ende, 2008), but it is unclear what factors determine differences in parental and young children's reports of problem behavior. Finally, many traditional studies relied on parent reported outcomes only, but as described parents' perspective may also be limited. The BPI may prove to be a valuable addition to existing methods. In general, the use of data from multiple informants (e.g. parents and children) will not result in complete agreement and inherently provides partly discrepant information. Yet, in some instances information can simply be summed to obtain accurate descriptions of child problems, whereas in other instances more sophisticated analytical approaches are needed to combine information. Sometimes it may prove important to compare results obtained using parent reported outcomes with results obtained using child reported outcomes of the same problems. This may help to examine whether results are consistent, and whether child information can aid in obtaining more accurate descriptions of child problem behavior.

Aim of this thesis

The main aim of the present thesis was to introduce the Berkeley Puppet Interview (BPI) in the Netherlands and to examine whether information about emotional, behavioral and peer relation problems, obtained by an interview with young children, would be valuable for use in epidemiological studies and eventually in clinical settings.

Specifically, the aims were to examine whether valid and reliable responses could be obtained about emotional, behavioral and peer relation problems from children aged five to eight years in a large general population sample. In addition, we studied whether children's information was consistently related to information from different sources like parent reports of problem behavior, indices of treatment referral, and DSM-based diagnoses. Further, we examined whether child self-report of problem behavior could be of additional value besides parent report of child problem behavior, for instance in studies of the association between parental behaviors like harsh discipline and parental depression and child problem behavior.

The Generation R Study

All studies described in this thesis were embedded in the Generation R Study. This is a population-based cohort examining early determinants of growth, health and development from fetal life into young adulthood (Jaddoe et al., 2012; Tiemeier et al., 2012). Pregnant women living in the area of Rotterdam, the Netherlands, with an expected delivery date between April 2002 and January 2006 were invited to participate. In total, 9,778 mothers were enrolled in the study and gave birth to 9,749 known live-born children. At enrolment, maternal, paternal and demographic characteristics

were collected. Throughout the prenatal and postnatal phase up till age nine, parents reported on the development of their child by questionnaire. All children followed unto age five (n = 8,305), were invited for detailed measurements during a research center visit. A number of 6,690 children visited the research center where the BPI was assessed.

Outline

In this thesis various aspects of young children's reports of emotional, behavioral and peer relation problems are discussed. The main focus of chapter 2 is the investigation of psychometric properties of the BPI in a large population-based sample. In this chapter the factor structure, the validity as indexed by associations with socio-demographic factors and the reliability of the BPI are examined. In chapter 3, results of the previous chapter are extended by examining the degree to which internalizing and externalizing problems, reported by the child on the BPI, are related to parent reported DSM-based diagnoses and treatment referral. Chapter 4 discusses the determinants of discrepancies in mother and child reports of child problem behavior. Discrepancies often arise when information from multiple informants is used to assess problem behavior. Yet, independent determinants of discrepancies in reports of mothers and young children are not well known. Chapters 5 and 6 address the value of using child reports in addition to reports of other informants. In chapter 5 associations of maternal and paternal depression with children's problem behavior are examined using information from mothers and young children. Results of different informants are compared to examine whether associations between parental depression and child problems are inflated if parents report on both their own depression and child problems. Chapter 6 examines the independent contribution of young children's reports above parental report of emotional and behavioral problems in a study of maternal and paternal harsh parenting. Finally, in chapter 7, parental harsh and sensitive parenting were examined in relation to child reported BPI peer aggression and peer relationship problems. In addition, the mediating role of children's self-regulation abilities in this relation is examined. The concluding chapter of this thesis, chapter 8, discusses the main findings of the studies described in this thesis, including general methodological considerations and implications for research and practice.

REFERENCES

Ablow, J.C., & Measelle, J.R. (2003). *Manual for the Berkeley Puppet Interview: Symptomatology, Social, and Academic Modules (BPI 1.0)*. MacArthur foundation Research Network Psychopathology and Development (David J. Kupfer, Chair): The MacArthur Working Group on Outcome Assessment

Ablow, J.C., Measelle, J.R., Cowan, P.A., & Cowan, C.P. (2009). Linking marital conflict and children's adjustment: The role of young children's perceptions. *Journal of Family Psychology, 23*(4), 485-499.

Ablow, J.C., Measelle, J.R., Kraemer, H.C., Harrington, R., Luby, J., Smider, N., et al. (1999). The MacArthur Three-City Outcome Study: Evaluating Multi-Informant Measures of Young Children's Symptomatology. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(12), 1580-1590.

Achenbach, T., Edelbrock, C., & Howell, C. (1987a). Empirically based assessment of the behavioral/emotional problems of 2- and 3- year-old children. *Journal of Abnormal Child Psychology, 15*(4), 629-650.

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987b). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213-232.

Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffitt, T. (2005). Psychometric Evaluation of 5- and 7-Year-Old Children's Self-Reports of Conduct Problems. *Journal of Abnormal Child Psychology, 33*(5), 537-550.

Baumeister, R.F. (2010). The self. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of science* (pp. 139-175): Oxford University Press.

Bruck, M., Ceci, S.J., & Hembrooke, H. (2002). The nature of children's true and false narratives. *Developmental Review, 22*(3), 520-554.

Ceci, S.J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*(3), 403-439.

De Los Reyes, A., & Kazdin, A.E. (2005). Informant Discrepancies in the Assessment of Childhood Psychopathology: A Critical Review, Theoretical Framework, and Recommendations for Further Study. *Psychological Bulletin, 131*(4), 483-509.

Eccles, J., Wigfield, A., Harold, R.D., & Blumenfeld, P. (1993). Age and Gender Differences in Children's Self- and Task Perceptions during Elementary School. *Child Development, 64*(3), 830-847.

Edelbrock, C., Costello, A.J., Dulcan, M.K., Kalas, R., & Conover, N.C. (1985). Age Differences in the Reliability of the Psychiatric Interview of the Child. *Child Development, 56*(1), 265-275.

Eder, R.A. (1989). The emergent personologist: The structure and content of 3½-, 5½-, and 7½-year-olds' concepts of themselves and other persons. *Child Development*, 1218-1228.

Eder, R.A. (1990). Uncovering Young Children's Psychological Selves: Individual and Developmental Differences. *Child Development, 61*(3), 849-863.

Goodvin, R., Meyer, S., Thompson, R.A., & Hayes, R. (2008). Self-understanding in early childhood: associations with child attachment security and maternal negative affect. *Attachment and Human Development, 10*(4), 433-450.

Grills, A.E., & Ollendick, T.H. (2002). Issues in Parent-Child Agreement: The Case of Structured Diagnostic Interviews. *Clinical Child and Family Psychology Review, 5*(1), 57-83.

Irwin, E.C. (1985). Puppets in therapy: An assessment procedure. *American Journal of Psychotherapy, 39*(3), 389-400.

Jaddoe, V.W., van Duijn, C.M., Franco, O.H., van der Heijden, A.J., van Iizendoorn, M.H., de Jongste, J.C., et al. (2012). The Generation R Study: design and cohort update 2012. *European Journal of Epidemiology, 27*(9), 739-756.

Kraemer, H.C., Measelle, J.R., Ablow, J.C., Essex, M.J., Boyce, W.T., & Kupfer, D.J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*(9), 1566-1577.

Marsh, H.W., Craven, R.G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology, 83*(3), 377-392.

Marsh, H.W., Ellis, L.A., & Craven, R.G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology, 38*(3), 376-393.

Measelle, J.R., Ablow, J.C., Cowan, P.A., & Cowan, C.P. (1998). Assessing Young Children's Views of Their Academic, Social, and Emotional Lives: An Evaluation of the Self-Perception Scales of the Berkeley Puppet Interview. *Child Development, 69*(6), 1556-1576.

Rutter, M. (1989). Isle of Wight revisited: Twenty-five years of child psychiatric epidemiology. *Journal of the American Academy of Child & Adolescent Psychiatry, 28*(5), 633-653.

Rutter, M., & Stevenson, J. (2008). Developments in Child and Adolescent Psychiatry Over the Last 50 Years. In M. Rutter, D. V. M. Bishop, D. S. Pine, S. Scott, J. Stevenson, E. Taylor & A. Thapar (Eds.), *Rutter's Child and Adolescent Psychiatry, 5th Edition*: Blackwell Publishing Limited.

Tiemeier, H., Velders, F.P., Szekely, E., Roza, S.J., Dieleman, G., Jaddoe, V.W., et al. (2012). The Generation R Study: A review of design, findings to date, and a study of the 5-HTTLPR by environmental interaction from fetal life onward. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(11), 1119-1135 e1117.

Valla, J.P., Bergeron, L., & Smolla, N. (2000). The Dominic-R: a pictorial interview for 6- to 11-year-old children. *Journal of the American Academie of Child & Adolescent Psychiatry, 39*(1), 85-93.

van de Looij-Jansen, P.M. (2010). *Self-report in Youth Health Monitoring: evidence from the Rotterdam Youth Monitor*: Erasmus MC: University Medical Center Rotterdam.

Verhulst, F.C., & van der Ende, J. (2008). Using Rating Scales in a Clinical Context. In M. Rutter, D. V. M. Bishop, D. S. Pine, S. Scott, J. Stevenson, E. Taylor & A. Thapar (Eds.), *Rutter's Child and Adolescent Psychiatry, 5th Edition*: Blackwell Publishing Limited.

Wu, P., Hoven, C.W., Bird, H.R., Moore, R.E., Cohen, P., Alegria, M., et al. (1999). Depressive and disruptive disorders and mental health service utilization in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(9), 1081-1090; discussion 1090-1082.

chapter 2

Young Children's Self-Reported Emotional, Behavioral and Peer Problems: The Berkeley Puppet Interview

Ringoot, A. P., Jansen, P. W., Steenweg-de Graaff, J., Measelle, J. R., van der Ende, J., Raat, H., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2013)

Psychological Assessment. 25(4), 1273-1285

ABSTRACT

Adult observers are typically the only informants of emotional and behavioral problems in young children. Although additional information can be provided by child self-report, few validated, structured instruments are available to obtain self-report from young children. The Berkeley Puppet Interview (BPI) has been developed to obtain structured self-reports on multiple domains of mental health and social well-being. This study was the first to evaluate the psychometric properties of the BPI in a large sample. We studied 8 a priori scales of the interview in a Dutch community sample of 6,375 children ages 5–7 years. Using confirmatory factor analysis, we demonstrated adequate fit (Tucker–Lewis Index = .90; Comparative Fit Index = .90; Root-Mean-Square Error of Approximation = .03) of a multidimensional model with 50 items loading on 8 latent factors (Depression, Separation Anxiety, Overanxious, Oppositional Defiant, Overt Hostility, Conduct Problems, Bullied by Peers, and Peer Acceptance/Rejection). This model was invariant across gender. Children reported anxiety-related problems more frequently than depressive problems, behavioral problems, or difficulties in peer relations. Reliability analyses showed that 3 broadband scales designated as Internalizing, Externalizing, and Peer Relations were homogeneous constructs (*α*s = .68 – .79). Higher scores on most BPI scales were associated with lower maternal education, lower family income, and non-Western ethnicity. Boys reported more behavioral and peer relation problems, whereas girls reported more emotional problems. The findings indicate that young children from socioeconomically and demographically diverse backgrounds are capable of providing valid, multidimensional information on their emotional, behavioral, and peer relation problems using the BPI. The use of young children's self-report could be a promising addition to existing assessment tools.

INTRODUCTION

Emotional and behavioral problems occur as early as preschool age. It is estimated that approximately 7%–26% of all preschool children (2–5 years old) and about 13% of school-age children (6 –16 years old) experience an emotional or behavioral disorder (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; Egger & Angold, 2006; Wichstrøm et al., 2011). Problem behavior impairs both young children and their families, tends to persist, and can precede adult psychiatric disorders (Reef, van Meurs, Verhulst, & van der Ende, 2010). Thus, early identification of problem behavior in children is important.

Traditionally, young children were thought to be limited in their ability to report on their own problems (e.g., Edelbrock, Costello, Dulcan, Kalas, & Conover, 1985). Therefore, the assessment of psychopathology in young children, particularly in children younger than eight years old, has relied largely on adult reports (Luby, Belden, Sullivan, & Spitznagel, 2007; Measelle, John, Ablow, Cowan, & Cowan, 2005). However, to obtain a more complete picture of a child's functioning, assessments based solely on adult informants may not be sufficient. Parents, teachers, or other adult caregivers may not be aware of certain problems in children. For example, emotional problems in children may be less apparent to observers because they are generally less disruptive (Achenbach, McConaughy, & Howell, 1987; Wu et al., 1999). Furthermore, each informant provides limited information as a result of different relationships with the child, differing standards for judging a child's functioning, and, importantly, the situational specificity of a child's behavior (De Los Reyes & Kazdin, 2005; Verhulst & van der Ende, 2008). If only parents or teachers are considered as informants, the perspective on the child's problems is generally limited. Consequently, the accuracy of diagnostic information may be improved by including child report regarding his or her own problems (Kraemer et al., 2003).

Despite concerns that the developmental immaturity of young children would influence their ability to report accurately on their own problems, studies in the past decades have suggested that children between 4 and 8 years old have more differentiated concepts of themselves than was previously assumed (Marsh, Craven, & Debus, 1991; Marsh, Ellis, & Craven, 2002). Self-concept research has shown that young children are capable of giving multidimensional self-descriptions and that their descriptions of academic competence, achievement motivation, relationships with caregivers and peers, and social competence are consistent with both observed behavior and more distal measures such as social relationships and attachment security (Colwell & Lindsey, 2003; Eccles, Wigfield, Harold, & Blumenfeld, 1993; Goodvin, Meyer, Thompson, & Hayes, 2008; Harter & Pike, 1984; Ladd, 1990; Marsh et al., 2002; Sessa, Avenevoli, Steinberg, & Morris, 2001). Most of these studies investigated the dimensionality of self-concept or described self-report of concrete observable constructs such as physical appearance, social functioning, self-esteem, and performance in sports or school. However, the studies did not tap into issues such as self-report of emotional and behavioral problems.

Several instruments have been developed that address self-reported emotional or behavioral problems in young children, but many of these target only specific problems such as depression or worry (e.g., Martini, Strayhorn, & Puig-Antich, 1990; Zeman, Cassano, Suveg, & Shipman, 2010). Only

a few interviews that can be used in children younger than 8 years cover a broad range of emotional and behavioral problems. Most notable among these are the Diagnostic Interview for Children and Adolescents (DICA; Reich, 2000), the Children's Interview for Psychiatric Syndromes (ChIPS; Weller, Weller, Fristad, Rooney, & Schecter, 2000), the Pictorial Interview for Children and Adolescents (PICA–III–R; Ernst, Cookus, & Moravec, 2000), and the Dominic–R (Valla, Bergeron, & Smolla, 2000). Of these four, the DICA and the ChIPS target children from 6 to 18 years old. Although age appropriateness was considered by adapting wording and sentence length, the interview formats are similar to the methods that are typically used for adult assessments. This may not be suitable for young children, given their less advanced language abilities and shorter attention span (Arseneault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005; Measelle, Ablow, Cowan, & Cowan, 1998). Reich (2000) reported that children understood DICA items better after the items were combined with a pictorial instrument, emphasizing the limitation of using adult interview formats in young children. The two other instruments, the PICA–III–R and the Dominic–R, were developed with an effort to adapt the interview format to the cognitive ability of young children. These methods included the use of drawings to capture children's interest and to illustrate the interview questions. Yet, to the best of our knowledge, no further work on the PICA– III–R has been done since a validation study in a small clinical sample (Ernst et al., 2000). The Dominic–R has been studied more extensively, and it can provide reliable, standardized information from young children (Valla et al., 2000). However, none of the four interviews target children younger than the age of six. There is thus a need for reliable and valid structured child-interviews, assessing a broad range of emotional and behavioral problems that can be used in an age range including children younger than six years old.

The Berkeley Puppet Interview (BPI) was developed to assess emotional and behavioral problems, social functioning, and school functioning in children between the ages of four and eight (Ablow, Measelle, & the MacArthur Working Group on Outcome Assessment, 2003; Ablow et al., 1999; Measelle et al., 1998). The BPI was originally developed to assess self-concept in young children (Measelle et al., 1998). While the initial self-concept scales were derived empirically using principal component analysis, the scales in the present version of the BPI that assess emotional and behavioral problems were developed in parallel with the adult report Health and Behavior Questionnaire (HBQ; Ablow et al., 1999; Boyce et al., 2002), and items within the scales were derived mainly to map onto the system (Ablow et al., 2003) of the Diagnostic and Statistical Manual of Mental Disorders 4th edition (DSM-IV; American Psychiatric Association, 1994). The BPI was developed to account for several factors that might complicate self-report by young children, such as short attention span, less developed vocabulary, and age appropriateness (Measelle et al.,1998). The method was based on earlier studies that showed that the use of puppets keeps young children engaged (e.g., Eder,1990). Previous studies using the BPI have indicated that it is a feasible instrument to obtain children's reports of emotional and behavioral problems (Ablow et al., 1999; Arseneault et al., 2005; Luby et al., 2007), competencies, personality traits, and close interpersonal relationships (Ablow, Measelle, Cowan, & Cowan,2009; Measelle et al., 1998; 2005). Ablow et al. (1999) investigated the psychometric properties of the BPI scales that specifically target emotional and behavioral problems. They demonstrated good test–retest reliability and significant

discriminant validity of most scales. Factor analysis on scale scores yielded an externalizing and an internalizing factor. However, a factor analysis on item level was not done, and the study was performed only in a small, ethnically homogeneous sample of 120 children (67 community and 53 clinical children). In other studies examining the psychometric properties of the BPI, only a limited set of problem domains, such as depression, conduct problems and familial conflict, were explored (Ablow et al., 2009; Arseneault et al., 2005; Luby et al., 2007).

We examined the psychometric properties of the BPI in a large general population sample. Eight scales were selected that measure peer relations and emotional and behavioral problems common in young children. For comparability across studies, we used the original BPI scales that adhere to DSM nosology (Ablow et al., 2003) and performed a confirmatory factor analysis (CFA) to test whether a multi-dimensional model with eight factors would be supported. In addition, models with fewer dimensions were tested to study whether young children can distinguish emotional, behavioral, and peer relation problems. We also evaluated the score distribution and internal consistencies of the BPI scores. To further explore whether BPI scores can be interpreted validly, we examined cross-scale correlations, associations of the BPI scores with demographic and socioeconomic characteristics that are known correlates of problems in children, and we studied whether these associations were independent of maternal report of problem behavior. Finally, cross-informant correlations were examined.

METHOD

Design

This study was embedded in the Generation R Study, a population based prospective cohort from fetal life onwards. The design and data collection methods have previously been described in detail (Jaddoe et al., 2012; Tiemeier et al., 2012). All children were born between April 2002 and January 2006 in the area of Rotterdam, the Netherlands. Typically, enrollment took place in early pregnancy. The study was conducted in accordance with the guideline proposed in the World Medical Association Declaration of Helsinki (World Medical Association, 2002) and has been approved by the Medical Ethics Committee of Erasmus Medical Center, Rotterdam. Written informed consent was obtained from all participants.

Population

From approximately the age of 5 years onward, participants were invited for a new study phase that started with the 6-year examination wave. Only participants with consent for the new study phase were eligible for the present study. Consent for participation in hands-on assessments or consent for participation with questionnaire data was obtained from caregivers of 8,305 children. Of the 8,305 participants who were eligible for follow-up examinations, 1,008 could not be reached for the 6-year examination wave and participated only with health care data linkage; 607 participants contributed questionnaire data only. This left 6,690 children and their caregiver; each child and

his or her caregiver visited our research center at the Erasmus Medical Center–Sophia Children's Hospital, for a variety of physical, behavioral, and biological assessments. Of the 6,690 children who visited the research center, 6,598 (99%) started the Berkeley Puppet Interview (BPI) during this visit (n = 92 children did not start the interview). We excluded data of children who were 8 years or older when the BPI was completed (n = 146), because we considered the BPI to be most suitable for children younger than 8 years of age. We also excluded child-interview data when the interviewer or the coder noted that the child did not understand the interview or if more than 25% of the data were missing on all scales (n = 77). This resulted in a total sample for analysis of 6,375 children from a multi-ethnic background: 3,181 boys and 3,194 girls, 5–7 years of age, with a mean age of 6.1 years (SD = 0.42). A graphical representation of the subject flow is presented in supplemental material (Figure S1).

In the Netherlands, most children attend school from the age of 4 years onward and attendance is compulsory from age 5. Thus, the majority of children had attended at least 1 year of Dutch schooling before participation.

Measures

Berkeley Puppet Interview.

Instrument. The Berkeley Puppet Interview (BPI; Ablow et al.,2003) is a semi-structured interactive interview technique to obtain self-reports from young children. The BPI consists of several interview modules and new modules are being developed continuously. At present, the BPI comprises more than 25 scales that assess emotional and behavioral problems, social well-being, academic functioning, and family environment (Ablow et al., 2003; 2009; Measelle et al., 2005). The BPI manual describes 17 scales in three domains: Symptomatology Scales (BPI–S), Social Scales (BPI–Soc), and Academic Scales (BPI–A). The BPI scales were constructed a priori, and items pertain to problems that are known to be important in young children (Ablow et al., 2003). Reliability and discriminant validity of the scales within the BPI-S, as well as the construction of these scales, have been described previously (Ablow et al., 1999; 2003). Users can select scales to assess the problems of their interest. Because of time constraints related to the large number of participants, we selected eight of the original scales that cover problems common in young children and translated these scales into Dutch. The translation procedure included forward and backward translation, expert panel discussion, and a pilot of the translation. The selection of eight scales resulted in a 50-item interview covering two Social Scales that assess Peer Relations (Bullied by Peers and Peer Acceptance/Rejection) and six Symptomatology Scales (Depression, Separation Anxiety, Overanxious, Oppositional Defiant, Overt Hostility and Conduct Problems). The first three scales target internalizing (emotional) problems, while the last three scales target externalizing (behavioral) problems (Ablow et al., 2003). Table 1 gives an overview of the scale items.

Procedure. During the interview, two identical dog hand puppets, named Iggy and Ziggy, were introduced to the child and invited him or her to engage in conversation. The puppets made opposing statements about themselves; for example, one puppet said, "I am a sad kid," while the other said, "I'm not a sad kid." Subsequently, the puppets asked children to indicate which statement

described themselves best: "How about you?" Questions were worded in age-appropriate language, and children could respond either verbally or nonverbally by pointing at the puppet. In this way, children became engaged in a natural dialogue with "peers". The BPI method does not use a forced-choice response format or recognition task. Children were allowed to respond in whatever way was comfortable to them. The interviewer would prompt in a standardized way only if a response was ambiguous. Positive and negative items were counterbalance between the two puppets within participants. The average time to complete the interview was 13.42 min (SD = 2.4; n = 1,753). The interviews were videotaped and coded. Children's responses were scored on a 7-point scale, with higher scores representing more problems. The exact score depended on which of the puppets' statements children chose and how much emphasis was put on the answer. For example, the actual negative or positive verbal or nonverbal response was coded with 6 ("I am a sad kid") or 2 ("I'm not a sad kid"). Amplification of an answer was coded 7 or 1 (e.g., "I always [or never] am a sad kid"). Code 5 or 3 was applied if a condition was added or if the answer was toned down ("usually," "most of the time"). Children's answers were coded 4 if a child indicated that the statement was true sometimes or if he or she agreed with both statements. Most answers received a code of either 2 or 6. There may have been some variation in the response patterns of children because some children were more talkative or more outgoing than others. However, this is unlikely to have caused true misclassification in coding the answers (i.e., a positive answer is scored as a negative answer or vice versa). All interviewers were research assistants who had been trained under the supervision of one of the developers of the BPI. Regular update sessions were conducted. Other research assistants coded the videotaped interviews independently of the interviewers. Instruction and supervision sessions were scheduled to train the coding procedure and to discuss encountered difficulties. Interrater reliability was assessed directly after the training and, to detect possible rater drift, was assessed repeatedly throughout the coding process. At the end of the training period, the average interrater reliability on the scale scores (intraclass correlation coefficients, ICCs) ranged from .87 (Overt Hostility) to .96 (Depression; n = 20). Throughout the coding process, the average interrater reliability for each scale ranged from .96 (Overanxious, Overt Hostility, Conduct Problems, and Bullied by Peers) to .98 (Depression, Separation Anxiety, Peer Acceptance/Rejection).

Additional measures

Information on maternal age, first born child (yes/no), maternal educational level, and ethnicity of the mother was obtained by questionnaire. Maternal education was defined as highest completed schooling and classified in three levels; "primary," "secondary" or "high" education. The mothers' marital status was categorized as "married or cohabiting" and "single parenthood". Family income, defined by the total net monthly income of the household, was categorized as less than 1,200 euro; 1,200–2,000 euro; or more than 2,000 euro. The mother's ethnicity was classified by the countries of birth of her parents, in accordance with the Dutch standard classification criteria of Statistics Netherlands (2004). A mother was classified as non-Dutch if one of her parents was born abroad.

Table 1. Item Prevalence and Standardized Parameter Estimates from the Eight-Factor Confirmatory Factor Analysis Model ($N = 6,345$)

Scale and Item[a]	Factor loadings	Prevalence (%)[b]		
		Problem present	Sometimes	Problem not present
Depression				
7. I'm a sad kid.	.53	7.3	15.0	77.7
12. I'm not a happy kid.	.54	5.1	4.6	90.3
14. I think I'm stupid.	.57	3.3	7.5	89.2
27. I don't like myself.	.51	6.3	3.2	90.5
34. I'm a lonely kid.	.51	10.5	3.3	86.2
41. I'm tired a lot.	.35	32.6	11.8	55.6
46. I cry a lot.	.37	23.2	12.2	64.6
Separation Anxiety				
4. I worry about Mom or Dad when I'm at school.	.72	14.2	4.3	81.5
9. I worry that Mom and Dad will go away and never come back.	.38	52.3	2.6	45.1
17. It's hard for me to say goodbye to my Mom or Dad.	.49	30.8	4.5	64.7
31. When I'm at school, I miss my Mom or Dad.	.60	21.2	6.3	72.5
36. I get scared if Mom or Dad goes somewhere without me.	.60	23.6	5.1	71.3
42. I mind going places without Mom or Dad.	.50	33.0	4.3	62.7
Overanxious				
1. I have lots of bad dreams.	.39	28.2	20.5	51.3
6. I get nervous when my teacher asks me a question.	.50	16.9	4.9	78.2
20. I get headaches a lot.	.47	22.5	15.2	62.3
22. I worry bad things are going to happen.	.47	34.5	5.2	60.3
24. If Mom or Dad isn't near my bed, I'm scared to go to sleep.	.54	17.5	5.6	76.9
26. I get tummy-aches a lot.	.49	30.0	16.2	53.8
33. I worry a lot.	.56	15.7	8.4	75.9
Oppositional Defiant				
29. When I don't get my own way, I yell at my teacher.	.73	3.2	0.8	96.0
32. When I get mad I lose my temper.	.39	36.6	8.3	55.1
39. I don't do what my parents ask me to do.	.40	14.0	13.9	72.1
43. When I don't get my own way, I yell at Mom or Dad.	.63	9.3	5.2	85.5
48. I swear or say bad words	.73	3.5	11.4	85.1
Overt Hostility				
5. When a kid gets in trouble at school, I think that's funny.	.55	8.8	3.6	87.6
11. I tease other kids.	.75	5.6	6.1	88.3
16. I hit other kids.	.78	4.9	8.5	86.6
19. I think it's fun to tease and pick on other kids.	.82	4.8	2.1	93.1
25. I like to mess up other kids' games or work.	.66	3.2	5.3	91.5
35. I often fight with other kids.	.72	11.2	10.7	78.1
38. If someone is mean to me, I hit them.	.73	13.5	4.2	82.3
Conduct Problems				
3. I take things that don't belong to me.	.37	10.5	27.1	62.4
8. When I want something that's not mine, I steal it.	.60	5.3	2.0	92.7
15. I tell lies.	.55	6.3	22.6	71.1
30. I'm mean to animals.	.56	2.5	5.9	91.6
40. I hit my Mom or Dad.	.52	3.7	14.6	81.7
44. I break other peoples' things when I'm mad at them.	.71	2.5	4.9	92.6
45. I start fires when a grown-up is not around.	.54	5.3	1.1	93.6
47. I hit my teacher or other grown-ups.	.72	1.0	1.7	97.3
49. When I'm playing a game, I cheat.	.56	6.1	18.1	75.8

Table 1. continued

Peer Acceptance/Rejection				
2. I don't have lots of friends at school.	0.32	20.1	9.8	70.1
10. Other kids don't ask me to play with them.	0.41	19.2	13.8	67.0
18. Kids don't like me.	0.77	10.6	11.9	77.5
23. Kids don't like to sit next to me.	0.58	11.0	8.4	80.6
50. When kids are playing games, they leave me out.	0.63	17.3	18.7	64.0
Bullied by Peers				
13. Kids do mean things to me.	0.66	22.3	22.8	54.9
21. Kids say mean things to me.	0.75	18.2	18.7	63.1
28. Kids at school hit me or beat me up.	0.74	9.5	12.7	77.8
37. Kids at school tease me.	0.72	15.6	13.9	70.5

Note. Data of subjects with > 25% missing data on all 50 items were not included in the analysis. [a]Numbers before the items indicate the order in the interview used for the current study. [b]We show prevalence of item scores aggregated to three categories, according to the structure that is present in the data: "problem present" (Codes 6 and 7), "problem a little or sometimes present" (Codes 3, 4 and 5), and "problem not present" (Codes 1 and 2).

Maternal ethnicity was categorized into Dutch, other-Western (i.e. other European, North American, and Oceanian), and non-Western (i.e. Turkish, Moroccan, Indonesian, Cape Verdean, Surinamese, and Antillean) national origin. The largest groups according to national origin in this sample were Dutch (56%), Turkish (8%), European (8%), Surinamese (7%), and Moroccan (5%). Information on the child's gender was derived from medical records. Child age at the visit to the research center was recorded.

Parent-reported emotional and behavioral problems were assessed using the Child Behavior Checklist for Ages 1½–5 (CBCL 1½–5; Achenbach & Rescorla, 2000), a 99-item questionnaire designed to obtain ratings of emotional and behavioral problems from parents about their children. Parents (in 93% of the cases, this was the mother) completed the questionnaire at home, prior to the research center visit. We used the CBCL for Ages 1½–5 for all children, as the majority (60%) of the children were younger than 6 years old at the time of CBCL assessment (37% were 6 years old and 3% of the children were 7 years old). We considered it important to use only one version of the CBCL during the same assessment wave to enhance comparability across all children. We chose to use the younger child version, as the version for 6to18-year-olds contains questions that are less applicable to 5-yearolds, such as questions concerning smoking tobacco. To study if items within the CBCL scales were rated equally consistent in the children younger than 6 and in children older than 6, we tested the internal consistencies of the scales in these two age groups and compared those. The internal consistencies for the Internalizing and Externalizing scales were similar in both age groups: in the younger age group, alphas were .85 and .90, respectively, and in the older age group, .87 and .90, respectively. This indicates that the scales were also reliably measured in older children. Parents rated the occurrence of their child's behavior within the past 2 months on a scale from 0 (not true) to 2 (very true or often true). The Internalizing scale score (36 items) was computed as the sum of the raw item scores of four syndrome scales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, and Withdrawn. The Externalizing scale score

(24 items) was derived by summing the raw item scores of the Attention Problems and Aggressive Behavior scales. Higher scores represent greater severity. Good reliability and validity have been reported for the CBCL (Achenbach & Rescorla, 2000).

Statistical Analysis

Feasibility. The feasibility of the interview was studied by the success rate (the fraction of children who completed the interview successfully) and the percentage of missing data on item level.

Score distribution. Score distribution was evaluated by assessing the prevalence of positive, negative, and neutral answers on each item, and by computing the mean scale scores of each BPI scale by summing the item responses.

Factor structure. Confirmatory factor analyses (CFAs) were conducted with Mplus Version 6.1 (Muthén & Muthén, 1998–2010). Our aim was to assess how well the 50 BPI items fitted the hypothesized eight-factor structure (Depression, Separation Anxiety, Overanxious, Oppositional Defiant, Overt Hostility, Conduct Problems, Bullied by Peers, and Peer Acceptance/Rejection). For this purpose, a multidimensional model with eight factors and 50 indicators was tested (Model 3). Three alternative models were also tested: a one-factor model (one-dimensional model; Model 1), a three-factor model representing three broad domains of Internalizing, Externalizing, and Peer Relation problems (Model 2), and a higher order model, with eight first-order factors and three higher order factors (Model 4). Since this study used Likert-type items, the items were treated as categorical (ordinal) rather than continuous data (Flora & Curran, 2004). To avoid possible distorting effects associated with low frequency cells, we performed CFAs on ratings of the original 7-point scale recoded into three groups, according to the structure that was present in the data. Responses on most items follow an M-shaped pattern: a low number of codes 1, 3, 4, 5, and 7 and a higher number of codes 2 and 6. To illustrate: for the first item in our interview, "I have a lot of bad dreams," the response percentages for each code were 2.0% for code 1; 49.3% for code 2; 12.3% for code 3; 2.1% for code 4; 6.0% for code 5; 25.8% for code 6; and 2.4% for code 7. Responses were recoded as follows: "problem not present" (original codes 1 and 2); "problem a little or sometimes present" (codes 3, 4, and 5); and "problem present" (codes 6 and 7*)*. We used a robust weighted least squares procedure with adjusted means and variance estimation, which allows for nonnormal data (Flora & Curran, 2004). Three indices that are relatively insensitive for sample size were applied to assess the goodness of fit of the models: comparative fit index (CFI), Tucker–Lewis index (TLI), and root-mean-square error of approximation (RMSEA). Good model fit was achieved if the TLI and CFI were equal to or greater than .95; a cutoff of .90 for the CFI and TLI was assumed to be reasonable. For RMSEA, a cutoff score of .06 or lower was implemented as criterion for goodness of fit (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004).

Measurement invariance. To determine the extent to which items have equal meanings across boys and girls, we tested a multigroup CFA with respect to gender in the multidimensional model (Model 3). The measurement invariance analysis was performed in two steps. First, a model without equality constraints was tested. Second, constraints on thresholds and factor loadings were added to the model. For categorical outcomes, measurement invariance models constrain thresholds and

factor loadings in tandem, resulting in this two-step procedure (Muthén & Muthén, 1998–2010). A CFI difference (Δ CFI) of -.01 or less was used to evaluate measurement invariance (Cheung & Rensvold, 2002).

Reliability. Internal consistency of the scales scores was estimated using Cronbach's alphas and corrected item-total correlations were inspected.

Cross-scale correlations. Correlations between scale scores were calculated with Pearson correlation coefficients.

Associations with demographic and socioeconomic characteristics. We examined the associations between BPI scores and demographic/socioeconomic correlates of child emotional and behavioral problems, including gender, child age, being a first-born child, maternal ethnicity, educational level of the mother, and marital status of mother. These associations were studied using analyses of variance (ANOVAs) and analyses of covariance (ANCOVAs). We first studied the association for each characteristic separately. Next, because of interrelatedness of the demographic and socioeconomic characteristics, we also studied the associations while controlling for the other characteristics by adding them as covariate to the analyses. In a final step, we examined whether these associations were independent of maternal report of problem behavior. To this aim, we repeated the ANCOVAs but additionally adjusted the analyses for mother reported problems on the CBCL. We used the broadband Internalizing and Externalizing Scales of the BPI and the CBCL, because these broadband scales target similar spectrums of problems. Percentages of missing values on the demographic and socioeconomic characteristics ranged from 0% for age and gender to 19% for income (mean percentage of missing data on all demographic and socioeconomic variables 6%). A dummy category for missing values was included in the ANCOVAs.

Cross-informant correlations. Cross-informant correlations were calculated between children's self-report on BPI Internalizing and Externalizing Scales and mother report of problems in similar broadband dimensions of the CBCL (Achenbach & Rescorla, 2000). These analyses were performed using SPSS Version 17 for Windows.

Response Analysis

To examine whether nonparticipation in the study was selective, demographic and socioeconomic characteristics of children who started the BPI (n = 6,598), were compared with data of children who were eligible but did not start the interview (n = 1,707). The group of children without BPI data more often had mothers of non-Western origin (43% vs. 36%), $\chi^2(1) = 24.83$, $p < .001$; had single parents (16% vs. 12%), $\chi^2(1) = 16.95$, $p < .001$; and more often came from families with lower maternal education (62% vs. 53%), $\chi^2(1)$ 36.61, $p < .001$, and lower family income (41% vs. 35%), $\chi^2(1) = 16.18$, $p < .001$, than children who were included in the study.

RESULTS

Feasibility

Of the 6,690 children who visited the research center, 6,598 (99%) started the interview. Of those who started, 77 children did not complete the interview successfully because they either had more than 25% missing data on all scales or did not understand the interview. Failure to start (n = 92) or complete (n = 77) the interview was due to: time constraints (n = 24); language/comprehension difficulties (n = 21); reluctance to participate (n = 18); technical issues (n = 17); severe anxiety (n = 16); problems related to the autism spectrum (n = 11); cognitive problems (n = 9); and other reasons including non-consent to be filmed, severe distractibility of the child, inadequate interview technique, or no documented reason for not completing the interview (n = 53). The mean percentage of missing data on item level was 1.5%, and ranged from 0.4% (Item 14) to 2.6% (Item 42).

Score Distribution

The endorsement of problems on each BPI-item is presented in Table 1. In general, children reported few problems. Some items, such as "I hit my teacher or other grown-ups," were particularly infrequent (problem present; 1%). On the contrary, other emotional and behavioral problems, particularly anxiety-related issues, appeared to be relatively common among children of this age. For example, 52% of the children reported worrying that their parents would leave without coming back, 34% worried that bad things were going to happen, and 33% said they minded going places without their parents.

Factor Structure

We performed four CFAs: a one-dimensional, a three-dimensional, a multidimensional and a higher order model. Table 2 presents fit indices associated with each model. Model 1 corresponds to a one-dimensional model testing a global index of problem behavior. This model did not meet the criteria specified for acceptable fit (CFI = .73, TLI = .72, RMSEA = .05). Model 2 comprised three factors that correspond with the broadband domains of internalizing, externalizing, and peer relation problems. Model 2 also did not meet the predefined cutoff for good model fit (CFI = .87, TLI = .86, RMSEA = .03). Model 3, the multidimensional model, comprised eight factors and demonstrated acceptable model fit (CFI = .90, TLI = .90, RMSEA = .03). Finally, although the fourth and higher order model did not reach the threshold for acceptable model fit, the fit indices were similar to those of Model 3 (Model 4: CFI = .89, TLI = .89, RMSEA = .03).

Table 1 shows standardized factor loadings, based on Model 3, for each item within its designated scale. All factor loadings exceeded .32 and had a p-value below .001. Factor loadings were lower in the internalizing spectrum and for peer relations than in the externalizing spectrum (ranges: Internalizing .35–.72, Peer Relations .32–.77, Externalizing .37–.82). Some factor loadings were relatively low, and the model fit might have been improved slightly by changes in the scale structure. However, because our approach was confirmatory rather than exploratory, and because

of the small number of items and the theoretical value of items within each scale, we decided not to use modification indices or to drop items with low factor loadings.

Measurement invariance. To test the equality of the model in boys and girls, we tested two models with a multigroup structure. The model without equality constraints had an acceptable fit to the data, $\chi 2$ (2,336) = 7,928, CFI = .91, TLI = .90, RMSEA = .03, as had the more restrictive model with constraints on thresholds and factor loadings, $\chi 2$ (2,378) = 7,534, CFI = .91, TLI = .91, RMSEA = .03. The difference in CFI (ΔCFI < -.01) indicated that the BPI was invariant across gender.

Table 2. Goodness-of-Fit Indices from Confirmatory Factor Analyses (*N* = 6,345)

Fit indices for the one-dimensional model versus multidimensional models	WLSMV $\chi 2$	*df*	CFI	TLI	RMSEA
Model 1: 50 items, 1 factor	18,459	1,175	.73	.72	.05
Model 2: 50 items, 3 factors	9,798	1,172	.87	.86	.03
Model 3: 50 items, 8 factors	7,269	1,147	**.90**	**.90**	**.03**
Model 4: 50 items, 8 factors, 3 second-order factors	8,174	1,164	.89	.89	.03

Note. Model 2: Three factors - Internalizing, Externalizing, and Peer Relations. Model 3: Eight factors - Depression, Separation Anxiety, Overanxious, Oppositional Defiant, Overt Hostility, Conduct Problems, Peer Acceptance/Rejection, and Bullied by Peers. Model 4: Higher order model with eight first-order factors and three second-order factors; second-order factors -Internalizing, Externalizing, and Peer Relations. Data of subjects with>25% missing data on all 50 items were not included in the analysis. Bold typeface indicates the most favorable model. WLSMV = weighted least squares adjusted for means and variances; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root-Mean-Square Error of Approximation.

Internal Consistency Reliability

Cronbach's alphas for the scores of each scale are presented in Table 3. The internal consistencies for global problem domains indicated acceptable reliability: Cronbach's alpha > .65 with alphas ranging from .68 (Peer Relations), through .72 (Internalizing), to .79 (Externalizing). When the eight scales were tested individually, only the scales for Hostility and Being Bullied reached acceptable levels of consistency. Corrected item-total correlations proved to be lower for emotional problems than for behavioral problems (mean Internalizing .28; mean Externalizing .36).

Cross-Scale Correlations

The item responses were summed to compute scale scores. Table 3 presents descriptive statistics. Pearson correlation coefficients for the BPI scale scores ranged from $r = .07$ to $r = .53$ (mean $r = .25$). Within the domains Internalizing, Externalizing and Peer Relations, the correlations between the scale scores were moderate: correlations were highest between the scales Overanxious and Depression/Separation Anxiety ($r = .42$) and between Overt Hostility and Conduct Problems ($r = .53$). Across domains, correlations between the scores on each of the eight scales were relatively small ($rs = .08 – .37$). Only the Depression Scale did not follow this pattern: all correlations between the depression and behavioral problem scale scores were above .22.

Table 3. Scale Descriptives, Cronbach's α and Correlations between Scale Scores.

Scale[a]	*N* items	Mean (*SD*)	100% range	Cronbach's *α*	Cross-scale correlations[b]										
					1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*A*	*B*	*C*
A. Internalizing	20	59.85 (12.8)	31.0–115.0	.72	.69	.77	.83	.18	.25	.25	.23	.36	—		
B. Externalizing	21	52.37 (11.2)	30.0–127.0	.79	.32	.11	.26	.73	.82	.85	.25	.32	.31	—	
C. Peer Relations	9	25.97 (7.5)	11.0–57.0	.68	.29	.15	.29	.23	.27	.24	.89	.66	.31	.29	—
1. Depression	7	18.50 (4.5)	11.0–41.0	.44	—										
2. Separation Anxiety	6	19.51 (6.0)	8.0–39.0	.57	.28	—									
3. Overanxious	7	21.84 (6.2)	9.0–44.0	.54	.42	.42	—								
4. Oppositional Defiant	5	13.43 (3.8)	5.0–31.0	.46	.22	.07	.15	—							
5. Overt Hostility	7	16.73 (4.9)	9.0–44.0	.73	.27	.10	.23	.39	—						
6. Conduct Problems	8	22.22 (5.1)	11.0–61.0	.56	.28	.09	.24	.46	.53	—					
7. Peer Acceptance/Rejection	5	14.18 (4.6)	6.0–32.0	.53	.22	.11	.20	.20	.21	.19	—				
8. Bullied by Peers	4	11.92 (4.5)	5.0–28.0	.69	.32	.15	.37	.22	.29	.25	.33	—			

[a]The n varies between scales since participants with >25% missing data on a subscale were excluded from the analysis for the particular subscale. No. of participants tested for each scale: Depression (*n* = 6,332), Separation Anxiety (*n* = 6,322), Overanxious (*n* = 6,333), Oppositional Defiant (*n* = 6,328), Overt Hostility (*n* = 6,343), Conduct Problems (*n* = 6,328), Peer Acceptance/Rejection (*n* = 6,347), and Bullied by Peers (*n* = 6,344). [b]All correlations were significant at the $p \leq .001$ level.

Associations With Demographic and Socioeconomic Characteristics

Table 4 presents fully adjusted associations of the BPI scale scores with demographic and socioeconomic characteristics. Unadjusted and adjusted analyses showed largely similar results, although some effects of birth order (first born or not), educational level, ethnicity, and marital status were more pronounced in unadjusted analyses (data of unadjusted analyses not shown). We present and discuss the results of the adjusted analyses. These analyses showed that boys and girls scored differently on the BPI: boys reported more behavioral problems and more negative peer relations, whereas girls reported more emotional problems. Gender explained most of the variance in the scores of the externalizing scales. Further, lower maternal education and non-Western origin were each associated with higher levels of self-reported problems on most BPI scales with the exception of the Oppositional Defiant, Conduct Problems, and Peer Acceptance/Rejection Scales. Lower family income (<1,200 euro monthly) was associated with higher scores on all scales except the Oppositional Defiant Scale. Next, we examined whether the associations of demographic and socioeconomic characteristics with young child self-report on the BPI were independent of mother reported problem behavior on the CBCL. To test this, we additionally adjusted the associations between sociodemographic characteristics and child report for mother reported problem behavior. These analyses, presented in Table 5, demonstrated that although the strength of nearly all associations was reduced, the associations between demographic and socioeconomic characteristics with BPI report remained and were largely independent of maternal report. In other words, the associations could not be explained with information obtained from the mother only

Cross-Informant Correlations

The child self-report data obtained with the BPI during the research center visit was compared with parent report on the mailed CBCL questionnaire. The median time interval between BPI and CBCL assessment was 1.05 months (90% range of time interval: 0.03–3.89 months). Item and scale content of the BPI and CBCL are not identical, but the broadband Internalizing and Externalizing Scales target the same spectrum of problems. Therefore, we computed Pearson's correlation coefficients on CBCL and BPI broadband scale scores. The correlations between child self-report data obtained with the BPI and parent report on the CBCL questionnaire were small, but significant. Correlations between BPI and CBCL scores within the internalizing and externalizing domains were higher than across the internalizing and externalizing domains: BPI and CBCL Internalizing Scale $r = .14, p < .001$; BPI and CBCL Externalizing Scale $r = .19, p < .001$; BPI Internalizing Scale and CBCL Externalizing Scale $r = .11, p < .001$; BPI Externalizing and CBCL Internalizing $r = .08, p < .001$.

Table 4. Associations Between Demographic and Socioeconomic Characteristics and Berkeley Puppet Interview Scale Scores

		Internalizing[a]			Externalizing[a]			Peer Relations[a]	
		Depression	Separation Anxiety	Overanxious	Oppositional Defiant	Overt Hostility	Conduct Problems	Peer cceptance/ Rejection	Bullied by Peers
Variable	*n*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*
Child characteristics									
Gender									
Boys	3,181	18.30 (0.08)	18.96 (0.10)	21.55 (0.11)	13.86 (0.07)	17.89 (0.08)	23.02 (0.09)	14.52 (0.08)	12.03 (0.08)
Girls	3,194	18.69 (0.08)***	20.07 (0.10)***	22.12 (0.11)***	13.00 (0.07)***	15.58 (0.08)***	21.42 (0.09)***	13.84 (0.08)***	11.81 (0.08)
F(df)		12.71 (1)***	55.25 (1)***	14.61 (1)***	80.77 (1)***	378.94 (1)***	157.37 (1)***	35.32 (1)***	3.70 (1)
Age									
F(df)	6,375	0.19 (1)	26.15 (1)***	3.07 (1)	4.80 (1)*	1.10 (1)	0.08 (1)	2.84 (1)	0.78 (1)
Firstborn									
Yes	3,498	18.37 (0.08)	19.52 (0.10)	21.72 (0.10)	13.45 (0.07)	16.75 (0.08)	22.17 (0.09)	14.03 (0.08)	11.87 (0.08)
No	2,670	18.69 (0.09)*	19.53 (0.12)	22.02 (0.12)	13.36 (0.08)	16.66 (0.09)	22.25 (0.10)	14.36 (0.09)*	11.98 (0.09)
F(df)		4.52 (2)*	0.51 (2)	2.36 (2)	2.57 (2)	1.78 (2)	1.18 (2)	4.40 (2)*	0.56 (2)
Maternal characteristics									
Income (monthly, in euros)									
>2,000	3,429	18.14 (0.09)	19.14 (0.12)	21.31 (0.12)	13.43 (0.08)	16.30 (0.09)	22.04 (0.10)	13.80 (0.09)	11.67 (0.09)
>1,200–2,000	923	18.48 (0.15)	19.72 (0.20)	22.14 (0.20)**	13.47 (0.13)	16.84 (0.16)*	21.97 (0.17)	14.52 (0.15)***	12.02 (0.15)
<1,200	822	19.14 (0.18)***	19.98 (0.24)*	22.71 (0.24)***	13.43 (0.15)	17.45 (0.19)***	22.83 (0.20)**	14.78 (0.18)***	12.49 (0.18)**
F(df)		11.62 (3)***	5.90 (3)**	11.18 (3)***	0.05 (3)	12.68 (3)***	4.72 (3)**	9.95 (3)***	5.27 (3)**
Ethnicity of mother									
Dutch	3,345	18.02 (0.09)	18.97 (0.12)	20.95 (0.12)	13.49 (0.08)	16.45 (0.10)	22.20 (0.10)	14.02 (0.09)	11.65 (0.09)
Other-Western	495	18.40 (0.20)	19.82 (0.27)*	21.89 (0.27)**	13.54 (0.18)	16.73 (0.22)	22.31 (0.23)	14.26 (0.21)	11.87 (0.21)
Non-Western	2,088	19.02 (0.11)***	20.22 (0.15)***	23.16 (0.15)***	13.17 (0.10)	17.12 (0.12)***	22.26 (0.13)	14.26 (0.11)	12.19 (0.11)**
F(df)		16.72 (3)***	14.50 (3)***	42.34 (3)***	3.46 (3)*	6.23 (3)***	0.13 (3)	1.66 (3)	4.94 (3)**

Table 4. continued

Educational level									
High	2,768	18.37 (0.10)	19.20 (0.13)	21.41 (0.13)	13.57 (0.09)	16.42 (0.11)	22.23 (0.11)	14.09 (0.10)	11.62 (0.10)
Secondary	2,520	18.60 (0.10)	19.72 (0.13)*	21.97 (0.13)*	13.42 (0.08)	16.92 (0.10)**	22.23 (0.11)	14.27 (0.10)	12.18 (0.10)**
Primary	523	19.10 (0.21)*	19.52 (0.28)	22.54 (0.29)**	13.32 (0.18)	17.24 (0.23)**	22.39 (0.24)	14.61 (0.22)	12.50 (0.21)***
F(df)		3.67 (3)*	3.14 (3)*	5.80 (3)**	1.65 (3)	5.31 (3)**	0.28 (3)	1.93 (3)	7.45 (3)***
Marital status									
Married/cohabiting	5,128	18.52 (0.07)	19.56 (0.09)	21.88 (0.09)	13.42 (0.06)	16.75 (0.07)	22.23 (0.08)	14.20 (0.07)	11.91 (0.07)
No partner	684	18.72 (0.19)	19.75 (0.25)	21.97 (0.25)	13.51 (0.16)	16.70 (0.20)	22.00 (0.22)	14.15 (0.19)	12.08 (0.19)
F(df)		1.53 (2)	1.46 (2)	0.68 (2)	0.13 (2)	0.12 (2)	0.56 (2)	0.09 (2)	0.38 (2)

Note. For each variable, the reference group is always the first category. Each sociodemographic characteristic was adjusted for all other variables in the table. Values after each sociodemographic category represent adjusted means and standard errors. The overall *F* statistic and associated p values were derived from analyses of covariance (ANCOVAs). The *p* values for difference were derived from ANCOVAs with pairwise comparisons, with Bonferroni adjustment for multiple comparisons.

[a] The *n* varies between scales since participants with > 25% missing data on a subscale were excluded from the analysis for the particular subscale. No. of participants tested for each scale: Depression ($n = 6{,}332$), Separation Anxiety ($n = 6{,}322$), Overanxious ($n = 6{,}333$), Oppositional Defiant ($n = 6{,}328$), Overt Hostility ($n = 6{,}343$), Conduct Problems ($n = 6{,}328$), Peer Acceptance/Rejection ($n = 6{,}347$), and Bullied by Peers ($n = 6{,}344$). * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5. Associations between Demographic and Socioeconomic Characteristics and BPI Scale Scores, Additionally Adjusted for Mother Report

	Before including mother report			After including mother report		
Variable	*n*	BPI Internalizing[a] *M (SE)*	BPI Externalizing[a] *M (SE)*	*n*	BPI Internalizing[a] *M (SE)*	BPI Externalizing[a] *M (SE)*
Child characteristics						
Gender						
Boy	3,181	58.80 (0.22)	54.76 (0.19)	2,709	58.12 (0.23)	54.23 (0.21)
Girl	3,194	60.89 (0.22)	50.00 (0.19)	2,720	60.21 (0.23)	50.17 (0.20)
F(df)		46.41 (1)***	300.95 (1)***		40.67 (1)***	194.08 (3)***
Age						
F(df)	6,375	11.52**	0.18	5,429	11.85**	0.58
First born						
Yes	3,498	59.62 (0.21)	52.37 (0.19)	3,031	58.75 (0.22)	52.06 (0.19)
No	2,670	60.25 (0.24)	52.24 (0.21)	2,223	59.84 (0.26)	52.19 (0.23)
F(df)		2.84 (2)	2.41 (2)		5.82 (2)**	4.25 (2)*
Maternal characteristics						
Income						
>2,000	3,429	58.60 (0.24)	51.75 (0.22)	3,223	58.26 (0.24)	51.84 (0.21)
1,200 – 2,000	923	60.34 (0.41)**	52.27 (0.37)	786	59.46 (0.44)	52.22 (0.39)
<1,200	822	61.84 (0.49)***	53.75 (0.44)**	600	60.97 (0.57)***	53.46 (0.50)*
F(df)		15.61 (3)***	6.06 (3)***		11.13 (3)***	2.76 (3)*
Ethnicity Mother						
Dutch	3,345	57.95 (0.25)	52.16 (0.22)	3,131	57.78 (0.24)	51.99 (0.22)
Other-Western	495	60.05 (0.56)**	52.64 (0.50)	442	59.37 (0.58)	52.13 (0.51)
Non-Western	2,088	62.40 (0.31)***	52.51 (0.27)	1,583	61.55 (0.34)***	52.29 (0.30)
F(df)		40.80 (3)***	0.55 (3)		25.30 (3)***	0.75 (3)
Educational level						
High	2,768	58.97 (0.27)	52.22 (0.24)	2,606	58.44 (0.27)	52.23 (0.23)
Secondary	2,520	60.29 (0.26)**	52.56 (0.24)	2,112	59.56 (0.28)*	52.20 (0.25)
Primary	523	61.24 (0.59)**	52.93 (0.52)	357	61.17 (0.70)**	52.74 (0.61)
F(df)		8.80 (3)**	0.79 (3)		5.26 (3)**	0.47 (3)
Marital Status						
Married/Cohabiting	5,128	59.96 (0.19)	52.39 (0.17)	4,554	59.25 (0.19)	52.24 (0.17)
No Partner	684	60.46 (0.52)	52.22 (0.40)	506	59.81 (0.59)	52.05 (0.52)
F(df)		2.01 (2)	0.07 (2)		1.79 (2)	0.16 (2)

Note. For each variable, the reference group is always the first category. Each sociodemographic characteristic was adjusted for all other variables in the table. Values after each sociodemographic category represent adjusted means and standard errors. Values for the overall *F*-statistic and associated *p*-values were derived from analyses of covariance (ANCOVAs). The *p*-values for difference were derived from ANCOVAs with pairwise comparisons, with Bonferroni adjustment for multiple comparisons.

[a]The *n* varies between scales since participants with >25% missing data on a subscale were excluded from the analysis for the particular subscale. Before including maternal report, Internalizing $n = 6,360$ and Externalizing $n = 6,333$. After including maternal report, Internalizing $n = 5,417$ and Externalizing $n = 5,397$. * $p < .05$, ** $p < .01$, *** $p < .001$.

DISCUSSION

This study presents psychometric properties of eight scales of the BPI. These scales assess peer relations and emotional and behavioral problems that are common in young children. In this large multiethnic community sample, results from CFAs supported the hypothesis that the BPI has a multidimensional structure, as the models with more dimensions (a higher order model and an eight-factor model) showed better model fit. The model with eight factors had the most favorable fit and fitted equally well in boys and girls. The factors derived from this model were associated with known correlates of emotional and behavioral problems. We also found that children gave internally consistent responses within the internalizing, externalizing, and peer relations broadband domains. The high success rate (i.e., the fraction of children who completed the interview successfully) and the low percentage of missing data at item level suggest that it is feasible to use the BPI with young children.

The finding that a multidimensional structure accurately describes the BPI data implies that young children are able to distinguish between distinct dimensions of emotional, behavioral, and peer relation problems.

Despite the confirmation of a multidimensional structure, each of the scales had only modest Cronbach's alphas. There are several possible explanations for this finding. First, the items within the scales tap heterogeneous constructs. For example, the Overanxious Scale includes items that assess somatic complaints. The low corrected item-total correlations also indicate that there is heterogeneity within the scales. Second, the small number of items per scale may have affected the reliability estimates, as alpha also depends on the number of items within a scale (Cortina, 1993). Indeed, the internal consistencies for the three broadband scales -Internalizing, Externalizing and Peer Relations, which consist of more items- were within an acceptable range. Finally, some of the items in this general population sample had low base rates (e.g., hitting the teacher), whereas other items were endorsed by many children (e.g., worrying that parents will leave). If many children oppose or endorse one, but not other items, this may seem inconsistent and can contribute to lower alphas. In clinical samples, there is less inconsistency in the item distribution of the BPI, and alphas have been reported to be higher (Ablow et al., 1999).

The Depression scale in particular provided lower internal consistencies than the other scales. Reporting consistently on emotional problems such as depression might be difficult for children, as this requires introspective abilities. Young children may be better at reporting concrete and core aspects of behavior (Harter, 1990; Luby et al., 2007). Alternatively, depressive problems and other emotional problems may be less domain specific, but more a generic expression of disturbance. This interpretation is also supported by the high comorbidity rates of oppositional defiant disorder and depression, and the overlap between emotional and behavioral problems in children from the general population (Egger & Angold, 2006; Wichstrøm et al., 2011). We did not examine comorbidities in the present study, but we found roughly the same level of correlation between depression and each of the other scales. This indicates that the depression scale did not differentiate between

emotional and behavioral problems as much as other scales did. This suggests that depressive symptoms are not specific for one problem, but a general indicator of a lack of well-being.

The levels of internal consistency we found are similar to levels reported in other studies using the BPI in nonclinical samples (Ablow et al., 1999; 2003; Arseneault et al., 2005). The modest alphas suggest that, even if an age-appropriate instrument is used, it is difficult for young children to report consistently on distinct emotional and behavioral problems (e.g., depression). Despite this difficulty, however, children can distinguish between problems as the CFA in our study showed. We also obtained consistent reports within broader domains of internalizing, externalizing, and peer relation problems.

Children's responses on the BPI varied substantially, with children reporting anxiety more frequently than other problems. This high prevalence of anxiety-related problems is in line with studies showing that anxiety is among the most prevalent psychiatric conditions in childhood (Costello, Egger, & Angold, 2005). Separation anxiety is particularly prevalent and typical for childhood (Cartwright-Hatton, McNicol, & Doubleday, 2006). Given that adults may identify behavioral problems better than they do anxiety and depression (Wu et al., 1999), the high prevalence of emotional problems on the BPI is an important finding. Future studies with the BPI must determine whether such information obtained by child self-report is indeed of added value when compared with adult report.

Cross-scale correlations on the BPI were moderate. In contrast, there are relatively high correlations between emotional and behavioral problems when parents report on their children's symptoms. For instance, the mean correlation between the CBCL Internalizing and Externalizing Scales is .50 (Achenbach & Rescorla, 2000), whereas the mean correlation between similar BPI scales in the present study was .31. These moderate correlations may well suggest, in line with the CFA results, that children distinguish between problems. Low correlations might theoretically also occur if children randomly responded to the items. However, each scale had higher correlations with other scales within its designated domain than across domains, which cannot easily be reconciled with a random answering pattern.

The BPI scales were associated with known demographic and socioeconomic correlates of problem behavior. Some interesting patterns appeared in these associations. Boys consistently reported more behavioral problems than girls, while girls reported more emotional problems than boys. Although we had expected to find some gender differences in problem behavior, the finding with respect to emotional problems is remarkable, as studies in young children have typically found that gender differences in depression or anxiety only emerge later in childhood or adolescence (Bongers, Koot, van der Ende, & Verhulst, 2003; Egger & Angold, 2006; but see Ablow et al., 1999). However, the studies that did not show a gender difference in depressive or anxiety-related problems generally relied on reports from adults, who may be less aware of emotional problems in children. Nonetheless, the differences between boys and girls in emotional problems were relatively minor. Given the large sample size, even small gender differences resulted in significant associations.

Indicators of socioeconomic disadvantage and non-Western national origin were related to higher levels of self-reported problems, except for the Oppositional Defiant and -to a lesser

extent- Conduct Problems scales. A similar effect of ethnicity was also reported in a study from the Netherlands (Bengi-Arslan, Verhulst, Ende, & Erol, 1997). In that study, the authors examined Turkish immigrant children, one of the largest ethnic minority groups in our study. They found that according to parents, Turkish children had more problems than Dutch children, specifically more anxious/depressed problems. In addition, concurrent with studies that have shown an inverse association of education, employment, and material circumstances with anxiety and depression, but less clear associations for other disorders (Bijl, Ravelli, & van Zessen, 1998; Fryers, Melzer, & Jenkins, 2003), we found no clear association between Conduct Problems and Oppositional Defiant Scales and indicators of socioeconomic disadvantage or non-Western national origin. Unlike the previous researchers, however, we did find an association of Overt Hostility with these characteristics.

The Overt Hostility Scale is different from the other behavioral problem scales, as it targets mainly whether a child bullies or victimizes children. The scales Bullied by Peers and Overt Hostility, which both assess involvement in bullying either as victim or as perpetrator, were associated with family income, ethnicity, and maternal education. Previous studies also have shown that immigrant children and ethnic minority groups are more likely to be involved in bullying (Strohmeier, Karna, & Salmivalli, 2011; Vervoort, Scholte, & Overbeek, 2010). Similarly, children of parents with a lower educational level and a lower family income are more likely to be bullied (Due et al., 2009; Nordhagen, Nielsen, Stigum, & Köhler, 2005; von Rueden et al., 2006). This means that our finding of socioeconomic and ethnic disparities in emotional and behavioral problems is consistent with previous literature and further supports the valid interpretation of the BPI results.

Many factors may have contributed to the low agreement between children's self-reports and maternal reports. For example, low agreement may in part be explained by the different assessment methods -puppet interview versus questionnaire- and by the different scales, items, and scoring methods that were used to compare mother reports and child reports. In addition, limited adult-child agreement may reflect the tendency of children, and possibly also parents, in the general population to report transient symptoms. There is evidence that if children's symptoms are more stable or more severe, cross-informant agreement is greater (Ablow et al., 2009). Furthermore, low agreement typically occurs as a result of variable settings, characteristics, and perspectives that influence parents' and children's reports of problem behavior (De Los Reyes & Kazdin, 2005). Low agreement could indicate that information from children on the BPI cannot be considered valid. Alternatively, low agreement could indicate that children and mothers provide unique information. Our results support the latter interpretation: known predictors of problem behavior like a lower family income and lower maternal educational level were associated with child-reported problem behavior, even after mother-reported problems were controlled. This indicates that children provide a unique perspective on their emotional and behavioral problems, independent of their mothers' reports of these behaviors. Thus, even though collecting child self-report may be time consuming and expensive, it is worthwhile. Child self-report may particularly be important in instances when nonresponse from other sources is high or when information about emotional problems is required, since this is more difficult to capture by an observer. In those instances, even young children can be considered a source of information in a multi-informant approach.

Our study has considerable strengths, including the large sample size and the socioeconomically diverse sample. Nevertheless, five main issues still need to be discussed. First, as the sample was population based, results cannot easily be generalized to clinical samples. Although psychometric properties tend to be more pronounced in less healthy samples, the feasibility of the BPI in a clinical setting and the extent to which the BPI can identify clinically relevant problems remain to be studied. Second, no test-retest reliability was assessed because children visited the research center only once, and we did not obtain approval for a second visit for test-retest purposes only. Previous studies using the BPI in community samples, have demonstrated moderate to high test–retest reliability that ranged from .43 to .86 on the scales used in the present study (Ablow et al., 1999; 2003). Third, we cannot rule out that coding of the responses on the BPI was influenced by aspects of children's personality or temperament, such as being more talkative and outgoing, or being more introverted or impaired in social communication. Future work may focus on developing rating scales of child behavior during the interview that include, and if necessary correct for, observed information in the analyses as well. Fourth, we assessed validity only by means of associations with demographic and socioeconomic characteristics, and cross-informant agreement of the BPI was studied only by a correlation with mother report. Additional information may be provided by more detailed investigation of associations of BPI data with other assessment tools and other informants. Finally, it is important to emphasize that the BPI items or scales do not yield a diagnosis. Rather, this child interview gives an indication of problems that are relevant to the child and should always be considered in conjunction with diagnostic information obtained from additional sources. Although we have shown that the BPI provides information independent of mother report, more detailed investigation of the incremental validity of child report is necessary (Mash & Hunsley, 2005). Arseneault et al. (2005) showed that child-reports using the BPI predicted disruptive classroom behavior two years later over and above mother report. The incremental validity of child report for other domains than conduct problems is worth studying.

Despite these limitations, our study indicates that young children can give multidimensional information on their emotional, behavioral, and peer relation problems. Their report of these common childhood problems can be interpreted validly. Children provide the most reliable information on broad domains of internalizing and externalizing problems and peer relations. Overall, young children's self-report using the BPI may provide a feasible addition to caregiver report in the assessment of emotional and behavioral problems.

REFERENCES

Ablow, J. C., Measelle, J. R., Cowan, P. A., & Cowan, C. P. (2009). Linking marital conflict and children's adjustment: The role of young children's perceptions. *Journal of Family Psychology, 23,* 485–499.

Ablow, J. C., Measelle, J. R., Kraemer, H. C., Harrington, R., Luby, J., Smider, N., . . . Kupfer, D. J. (1999). The MacArthur Three-City Outcome Study: Evaluating multi-informant measures of young children's symptomatology. *Journal of the American Academy of Child & Adolescent Psychiatry, 38,* 1580–1590.

Ablow, J. C., Measelle, J. R., & the MacArthur Working Group on Outcome Assessment. (2003). *Manual for the Berkeley Puppet Interview: Symptomatology, social, and academic modules (BPI 1.0)*. Pittsburgh, PA: University of Pittsburgh.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/ adolescent behavioral and emotional problems: Implications of crossinformant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington: University of Vermont, Research Center for Children, Youth & Families.

American Psychiatric Association. (1994). *The diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffitt, T. E. (2005). Psychometric evaluation of 5and 7-year-old children's self-reports of conduct problems. *Journal of Abnormal Child Psychology, 33,*537–550.

Bengi-Arslan, L., Verhulst, F. C., Ende, J., & Erol, N. (1997). Understanding childhood (problem) behaviors from a cultural perspective: Comparison of problem behaviors and competencies in Turkish immigrant, Turkish, and Dutch children. *Social Psychiatry and Psychiatric Epidemiology, 32,* 477-484.

Bijl, R. V., Ravelli, A., & van Zessen, G. (1998). Prevalence of psychiatric disorder in the general population: Results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Social Psychiatry and Psychiatric Epidemiology, 33,* 587–595.

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2003). The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology, 112,* 179–192.

Boyce, W. T., Essex, M. J., Woodward, H. R., Measelle, J. R., Ablow, J. C., & Kupfer, D. J. (2002). The confluence of mental, physical, social, and academic difficulties in middle childhood: Exploring the "headwaters" of early life morbidities. *Journal of the American Academy of Child & Adolescent Psychiatry, 41,* 580–587.

Cartwright-Hatton, S., McNicol, K., & Doubleday, E. (2006). Anxiety in a neglected population: Prevalence of anxiety disorders in pre-adolescent children. *Clinical Psychology Review, 26,* 817–833.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255.

Colwell, J. M., & Lindsey, E. W. (2003). Teacher–child interactions and preschool children's perceptions of self and peers. *Early Child Development and Care, 173,* 249–258.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98–104.

Costello, E., Egger, H., & Angold, A. (2005). 10-year research update review: The epidemiology of child and adolescent psychiatric disorders. I. Methods and public health burden. *Journal of the American Academy of Child & Adolescent Psychiatry, 44,* 972–986.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60,* 837–844.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483–509.

Due, P., Merlo, J., Harel-Fisch, Y., Damsgaard, M. T., Holstein, B. E., Hetland, J., . . . Lynch, J. (2009). Socioeconomic inequality in exposure to bullying during adolescence: A comparative, cross-sectional, multilevel study in 35 countries. *American Journal of Public Health, 99,* 907–914.

Eccles, J., Wigfield, A., Harold, R. D., & Blumenfeld, P. (1993). Age and gender differences in children's self and task perceptions during elementary school. *Child Development, 64,* 830 – 847.

Edelbrock, C., Costello, A. J., Dulcan, M. K., Kalas, R., & Conover, N. C. (1985). Age differences in the reliability of the psychiatric interview of the child. *Child Development, 56,* 265–275.

Eder, R. A. (1990). Uncovering young children's psychological selves: Individual and developmental differences. *Child Development, 61*.

Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry, 47,* 313–337.

Ernst, M., Cookus, B. A., & Moravec, B. C. (2000). Pictorial Instrument for Children and Adolescents (PICA–III–R). *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 94 –99.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466 – 491.

Fryers, T., Melzer, D., & Jenkins, R. (2003). Social inequalities and the common mental disorders. *Social Psychiatry and Psychiatric Epidemiology, 38,* 229 –237.

Goodvin, R., Meyer, S., Thompson, R. A., & Hayes, R. (2008). Self-understanding in early childhood: Associations with child attachment security and maternal negative affect. *Attachment & Human Development, 10,* 433– 450.

Harter, S. (1990). Developmental differences in the nature of self-representations: Implications for the understanding, assessment, and treatment of maladaptive behavior. *Cognitive Therapy and Research, 14.*

Harter, S., & Pike, R. (1984). The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children. *Child Development, 55*.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling 6,* 1–55.

Jaddoe, V. W., van Duijn, C. M., Franco, O. H., van der Heijden, A. J., van IJzendoorn, M. H., de Jongste, J. C., Hofman, A. (2012). The Generation R Study: Design and cohort update 2012. *European Journal of Epidemiology, 27,* 739 –756.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *The American Journal of Psychiatry, 160,* 1566 –1577.

Ladd, G. W. (1990). Having friends, keeping friends, making friends, and being liked by peers in the classroom: Predictors of children's early school adjustment? *Child Development, 61,* 1081–1100.

Luby, J. L., Belden, A., Sullivan, J., & Spitznagel, E. (2007). Preschoolers' contribution to their diagnosis of depression and anxiety: Uses and limitations of young child self-report of symptoms. *Child Psychiatry and Human Development, 38,* 321–338.

Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children 5 to 8 years of age: Measurement and multidimensional structure. *Journal of Educational Psychology, 83,* 377–392.

Marsh, H. W., Ellis, L. A., & Craven, R. G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology, 38,* 376–393.

Marsh, H. W., Hau, K. T., & Wen, Z. L. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler (1999) findings. *Structural Equation Modeling, 11,* 320–341.

Martini, D., Strayhorn, J. M., & Puig-Antich, J. (1990). A symptom self-report measure for preschool children. *Journal of the American Academy of Child & Adolescent Psychiatry, 29,* 594–600.

Mash, E. J., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology, 34,* 362–379.

Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing young children's views of their academic, social, and emotional lives: An evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development, 69,* 1556–1576.

Measelle, J. R., John, O. P., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (2005). Can children provide coherent, stable, and valid self-reports on the Big Five dimensions? A longitudinal study from ages 5 to 7. *Journal of Personality and Social Psychology, 89,* 90–106.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

Nordhagen, R., Nielsen, A., Stigum, H., & Köhler, L. (2005). Parental reported bullying among Nordic children: A population-based study. *Child: Care, Health and Development, 31,* 693–701.

Reef, J., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2010). Children's problems predict adults' *DSM–IV* disorders across 24 years.*Journal of the American Academy of Child & Adolescent Psychiatry, 49,*1117–1124.

Reich, W. (2000). Diagnostic Interview for Children and Adolescents (DICA). *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 59–66.

Sessa, F. M., Avenevoli, S., Steinberg, L., & Morris, A. S. (2001). Correspondence among informants on parenting: Preschool children, mothers, and observers. *Journal of Family Psychology, 15,* 53–68.

Statistics Netherlands. (2004). *Allochtonen in Nederland* 2004 [Foreigners in the Netherlands 2004]. Voorburg/ Heerlen: Centraal Bureau voor de Statistiek.

Strohmeier, D., Karna, A., & Salmivalli, C. (2011). Intrapersonal and interpersonal risk factors for peer victimization in immigrant youth in Finland. *Developmental Psychology, 47,* 248–258.

Tiemeier, H., Velders, F. P., Szekely, E., Roza, S. J., Dieleman, G., Jaddoe, V. W., . . . Verhulst, F. C. (2012). The Generation R Study: A review of design, findings to date, and a study of the 5-HTTLPR by environmental interaction from fetal life onward. *Journal of the American Academy of Child & Adolescent Psychiatry, 51,* 1119–1135.

Valla, J. P., Bergeron, L., & Smolla, N. (2000). The Dominic-R: A pictorial interview for 6to 11-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry, 39,* 85–93.

Verhulst, F. C., & van der Ende, J. (2008). Using rating scales in a clinical context. In M. Rutter, D. V. M. Bishop, D. S. Pine, S. Scott, J. Stevenson, E. Taylor & A. Thapar (Eds.), *Rutter's child and adolescent psychiatry* (5th ed.). Oxford, United Kingdom: Blackwell.

Vervoort, M., Scholte, R., & Overbeek, G. (2010). Bullying and victimization among adolescents: The role of ethnicity and ethnic composition of school class. *Journal of Youth and Adolescence, 39,* 1–11.

von Rueden, U., Gosch, A., Rajmil, L., Bisegger, C., Ravens-Sieberer, U., & the European KIDSCREEN Group. (2006). Socioeconomic determinants of health related quality of life in childhood and adolescence: Results from a European study. *Journal of Epidemiology and Community Health, 60,* 130–135.

Weller, E. B., Weller, R. A., Fristad, M. A., Rooney, M. T., & Schecter, J. (2000). Children's Interview for Psychiatric Syndromes (ChIPS). *Journal of the American Academy of Child & Adolescent Psychiatry, 39,*76–84.

Wichstrøm, L., Berg-Nielsen, T. S., Angold, A., Egger, H. L., Solheim, E., & Sveen, T. H. (2012). Prevalence of psychiatric disorders in preschoolers. *Journal of Child Psychology and Psychiatry, 53,* 695–705.

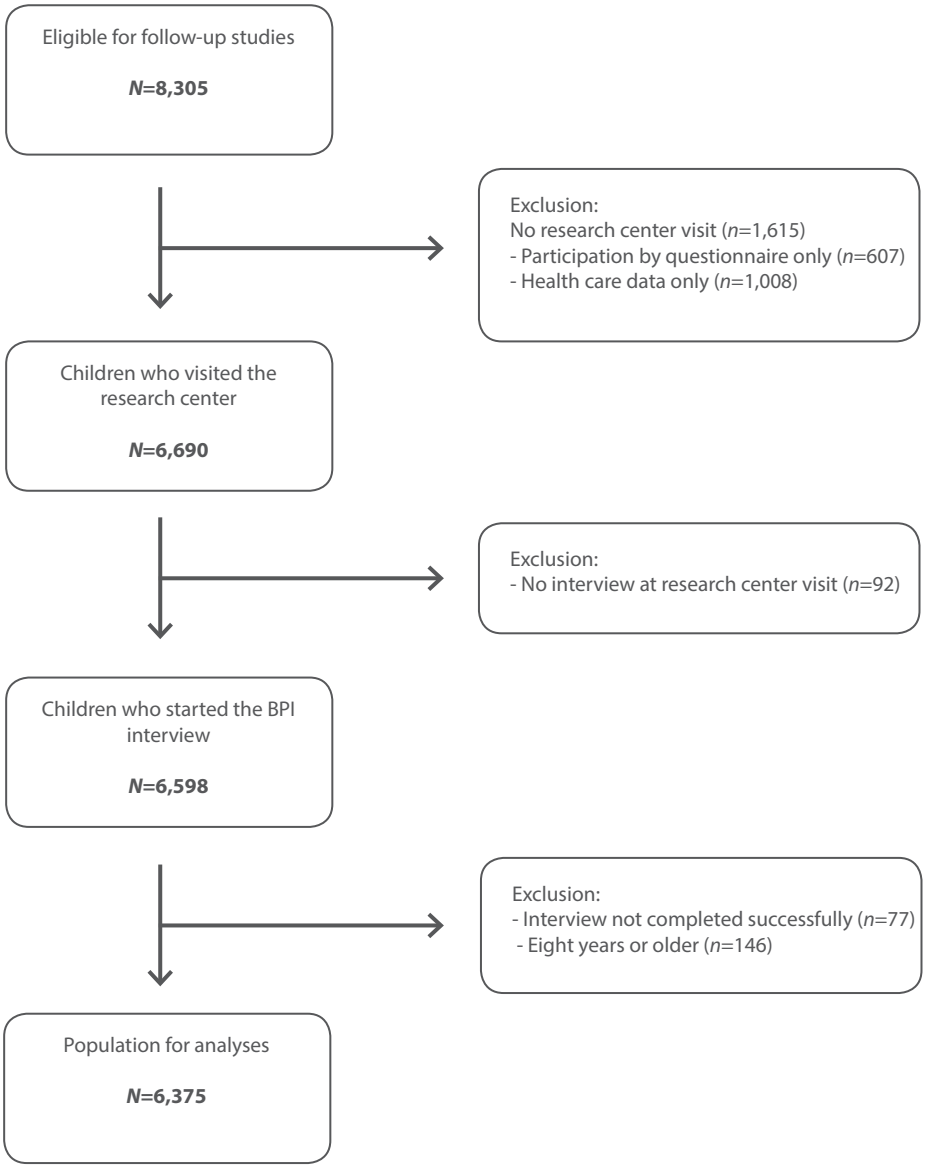
World Medical Association. (2002). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of Postgraduate Medicine, 48,* 206–208.

Wu, P., Hoven, C. W., Bird, H. R., Moore, R. E., Cohen, P., Alegria, M.,Roper, M. T. (1999). Depressive and disruptive disorders and mental health service utilization in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 38,* 1081–1090; discussion 1090–1082.

Zeman, J. L., Cassano, M., Suveg, C., & Shipman, K. (2010). Initial validation of the Children's Worry Management Scale. *Journal of Child and Family Studies, 19,* 381–392.

SUPPLEMENTARY MATERIAL

Figure S1 Participant Flow



chapter 3

Self-reported problem behavior in young children with and without a DSM-based diagnosis in the general population

Ringoot, A.P.; Jansen, P.W.; Rijlaarsdam, J.; So, P.;
Jaddoe, V.W.V.; Hofman, A.; Verhulst, F.C.; & Tiemeier, H.

Submitted for publication.

ABSTRACT

Problem behavior of young children is generally not assessed with structured child interviews. This paper examined how information about problem behavior, obtained by structured interviews with 6-year-old children, relates to DSM-based diagnoses obtained from parents and to treatment referral.

In a population-based cohort, parents of 1,084 children aged 5-8 years were interviewed with the DSM-based Diagnostic Interview Schedule-Young Child version (DISC-YC), and they scored the Child Behavior Checklist (CBCL). Children themselves were interviewed about problem behavior using the semi-structured Berkeley Puppet Interview (BPI). Treatment referral information was obtained by parent reported questionnaire.

Children with DSM-based diagnoses and children with CBCL problems in the clinical range reported more BPI problems. Associations were strongest in the externalizing domain (e.g. DISC-YC externalizing diagnoses with BPI externalizing scores: $F(1, 416) = 19.4$, $p < .001$; DISC-YC internalizing diagnoses with BPI internalizing scores: $F(1, 312) = 3.75$, $p = .54$). Higher BPI externalizing problem scores predicted treatment referral, although not independent of DISC-YC diagnoses obtained from parents.

Problems reported in a structured interview by young children are thus related to externalizing and –to a lesser extent– to internalizing DSM-based diagnoses, but this information has limited value in predicting treatment referral. Nevertheless, children's information may be useful to better understand problems relevant for the developing child.

INTRODUCTION

Accurate diagnosis of child psychopathology is of great importance for early and effective interventions, and can improve children's prognosis (Webster-Stratton & Taylor, 2001). Obtaining accurate diagnoses can, however, be challenging when psychopathology in young children is assessed. There is consensus that a multi-informant approach should be used to obtain a comprehensive picture of child problems (Kraemer et al., 2003). In the diagnostic process, clinicians and researchers generally rely on information from parents, observations of children's behavior and information from teachers or daycare professionals, but standardized information from the young children themselves is rarely obtained. Yet, adult informants may have difficulties recognizing children's problems, in particular if these are not very disruptive (Kolko & Kazdin, 1993; Wu et al., 1999). Hence, by not including structured child information, significant concerns young children have may be missed in the diagnostic process (Yeh & Weisz, 2001).

There is evidence suggesting that young children are capable of providing reliable, valuable and unique information about their emotions and behavior if age-appropriate, structured, instruments are used (Ablow et al., 1999; Valla, Bergeron, & Smolla, 2000). One such instrument is the Berkeley Puppet Interview (BPI; Ablow & Measelle, 2003), a semi-structured interview for children between four and eight years old. The interview uses puppets to engage young children in an interview to playfully discuss issues like depression, anxiety and aggression. The association of family socioeconomic status and teacher reports of school adjustment with children's information obtained with the BPI support the validity of this interview (Ablow et al., 1999; Arseneault et al., 2005; Ringoot et al., 2013). Although research also supports other psychometric properties of the BPI (e.g. Ablow et al., 1999; Ringoot et al., 2013), less is known about how children's reports on the interview relate to diagnostic information.

Only a few, relatively small, studies examined the association of young children's reports of, mostly selected, internalizing and externalizing problems with diagnostic information (Arseneault et al., 2005; Luby, Belden, Sullivan, & Spitznagel, 2007; Valla et al., 1994). There are no studies examining the relationship between child self-reported problem behavior, obtained with an interview suitable for children below the age of six years, and DSM-based diagnostic information across multiple problem domains. Some gaps in our knowledge thus remain.

The present paper aims to extend previous studies by examining the extent to which young elementary school children's internalizing and externalizing problems, reported in an interview with the child, are interrelated with child DSM-based diagnoses in a large population-based cohort. The Diagnostic Interview Schedule-Young Child version (DISC-YC) was used to determine parent reported DSM-IV-based diagnoses (Fisher & Lucas, 2006). We hypothesize that children with a DISC-YC derived DSM-based diagnosis report more problems than children without any diagnosis. To study consistency of findings across different measures, we also examined how information about internalizing and externalizing problems obtained during a child interview were related to scores in the clinical range on DSM-oriented scales of the Child Behavior Checklist 1½-5 (CBCL; Achenbach & Rescorla, 2000). Finally, because clinical practitioners not always use standardized diagnostic

interviews like the DISC-YC and because agreement between standardized diagnostic interviews and clinical evaluations is mostly modest (Rettew et al., 2009), we also examined whether higher BPI problems scores were associated with treatment referral.

MATERIALS AND METHODS

Design and Study Population

This study was embedded in the Generation R Study, a population-based cohort from fetal life onwards (Jaddoe et al., 2012). The Medical Ethics Committee at Erasmus MC, Rotterdam, the Netherlands approved the study. Written informed consent was obtained from all caregivers.

We used data obtained during the early school age (five to eight years) follow-up of the Generation R Study. Children and their caregivers filled out questionnaires and at a research center visit children were interviewed with the Berkeley Puppet Interview (BPI), 99% of the 6,690 children visiting the research center started the interview. After excluding poor quality BPI data and data of children who were older than 8 years at assessment, BPI data was available in a sample of 6,521 children (mean age 6.1 [*SD* = 0.4]).

To efficiently include children with potential DSM-IV disorders, children with high problem scores on the caregiver reported CBCL 1½-5, i.e. top 15% of total problem score and top 2% on any of the syndrome scores, were selected for in-depth diagnostic assessment with the DISC-YC. CBCL data was available in 6,172 children (mean age 6.0 [*SD* = 0.4], all younger than 8 years). Selected were 1,410 children: 1,080 children who scored above the cut-offs (screen-positive) and a random sample of 330 children who scored below the cut-offs (screen-negative). DISC-YC interview data was obtained from 1,154 children (82% of 1,410 selected). Eventually, analyses were performed in 1,084 children with available BPI, DISC-YC and CBCL data (77% of 1,410). In this sample, children were on average 6.7 years old at DISC-YC assessment (*SD* = 0.6, all younger than 9 years), 55% were boys, 63% were of Western national origin and 50% had a mother with a higher level of education. The average interval between BPI and DISC-YC assessment was 7.7 months (*SD* = 6.0), and 1.3 months (*SD* = 1.4) between BPI and CBCL assessment.

Questionnaire data on treatment referral was available in 746 of the 1,084 children with a mean age of 8.2 years (*SD* = 0.2), the average interval between assessment of the BPI and treatment referral was 25.4 months (*SD* = 3.9).

Measures

Child interview about internalizing and externalizing problems

Information on internalizing and externalizing problems as experienced by young children were assessed with the Berkeley Puppet Interview (BPI) (Ablow & Measelle, 2003), a semi-structured interview in which two identical dog hand puppets make opposing statements about themselves and ask the child to indicate which statement describes him/her best. Each statement was scored on a 7-point scale, ranging from 1 to 7. The exact score depended on which of the puppets' statements

the child chose and how much emphasis was put on the answer. Item scores were summed to compute scale scores. We used the 20-item broadband internalizing scale (α = .71) comprising the depression, separation anxiety and overanxious subscales, and the 21-item broadband externalizing scale (α = .78) comprising the conduct problems, hostility and oppositionality and defiant subscales. Within the Generation R Study, the BPI has shown an adequate factor structure, validity as indexed by associations with socio-demographic factors and acceptable internal consistencies of the internalizing and externalizing broadband scales (Ringoot et al., 2013).

Parent reported DSM-based diagnoses

The Diagnostic Interview Schedule for Children-Young Child version (DISC-YC; Fisher & Lucas, 2006), a structured and developmentally appropriate interview, was used to obtain DSM-IV based research diagnoses. Trained interviewers administered the computer assisted DISC-YC to caregivers during a home visit. For the present study, we used the diagnostic sections of mood disorder (dysthymia or major depression), separation anxiety, generalized anxiety disorder (GAD), specific phobia, oppositional defiant disorder (ODD), conduct disorder (CD) and attention deficit hyperactivity disorder (ADHD). Also, broad internalizing and externalizing diagnoses were defined (a diagnosis in the first four or latter three sections, respectively). Test-retest reliability of the DISC-YC symptom scales has previously been shown to be acceptable to high (see Lavigne et al., 2009).

At the end of each diagnostic section, the level of impairment from the symptoms was assessed using 6 questions. In the present study we report DISC-YC diagnoses if children displayed the minimum number of symptoms needed for a diagnosis and additionally experienced significant impairment in at least one area of functioning (defined as: symptoms "*often* prevented the child from doing things or going places" or "caused *very bad* problems with activities or with play"). These children were defined as 'cases'. All 'screen-negative' children without a DISC-YC diagnosis were defined as 'controls'.

Children with comorbid diagnoses on the DISC-YC were not studied separately since comorbid conditions are inherent to child psychopathology in the DSM-IV-system. There were 22 children with an internalizing and externalizing diagnosis.

Parent reported child internalizing and externalizing problems

The primary caregiver (92% mothers) completed the Child Behavior Checklist 1½-5 (CBCL 1½-5; Achenbach & Rescorla, 2000) at home. Caregivers rated the occurrence of their children's problem behavior on a 3-point scale ranging from 0 (not true) to 2 (very true or often true). Items were summed to compute scale scores with higher scores representing more problems. In the present study, we used the 36-item internalizing (α = .86) and 24-item externalizing (α = .90) broadband syndrome scales and the following DSM-oriented scales: affective problems, anxiety problems, oppositional defiant problems and attention deficit/hyperactivity problems (αs ranged between .53 for affective problems and .80 for attention deficit/hyperactivity problems). Children were classified as having problems if their score on the CBCL broadband or on the DSM-oriented scales was above the 91th and 98th percentile cut-off, respectively, of a Dutch norm group (Tick, van der

Ende, Koot, & Verhulst, 2007). Good reliability and validity have been reported previously for the CBCL (Achenbach & Rescorla, 2000).

Parent reported treatment referral.

Information on treatment referral was obtained by questionnaire filled in by parents when children were on average 8 years old. Parents indicated whether their child had received treatment for problem behavior in the past year, several examples of reasons for referral were provided. We aggregated all responses into one dichotomous variable: either referred for treatment or not.

Data analysis

Analyses were conducted in three steps. First, we examined whether children (cases) with a DISC-YC diagnosis reported more *BPI broadband externalizing and internalizing* problems than screen-negative children without any DISC-YC diagnosis (controls). This was examined using analyses of covariance (ANCOVAs) with child gender and age as covariates. We also studied whether gender modified any of these associations. To account for non-normality in the child interview data, the robustness of conclusions was tested by performing similar, but unadjusted, analyses using Mann-Whitney-U-tests; this did not meaningfully change findings and thus only results of parametric test are reported. We additionally used ROC analyses to test whether cut-offs on the externalizing and internalizing BPI scales could be established that optimally predict DISC-YC externalizing or internalizing diagnoses (Yesavage, 2002).

Second, to examine whether findings were similar across different assessment instruments or were specific to the DISC-YC, we tested whether children with a parent reported CBCL score in the clinical range reported higher levels of problem behavior on the child interview. These associations were again examined using ANCOVAs with child gender and child age as covariates.

Third, to test whether higher BPI problems scores predicted a clinical evaluation of psychopathology, logistic regressions were performed with BPI problem scores as determinant and treatment referral, on average two years later, as outcome. We also examined whether the prediction was independent of parental diagnostic information as obtained with the DISC-YC.

RESULTS

Table 1 presents children's mean externalizing and internalizing problem scores reported by the BPI, per diagnostic category of the DISC-YC. Children with any externalizing diagnosis reported more *externalizing problems* on the BPI ($M = 55.6$, $SE = 0.86$) than children without an externalizing diagnosis ($M = 50.8$, $SE = 0.68$; $F(1, 416) = 19.4$, $p < .001$, partial $\eta^2 = .05$). Likewise, children with specific externalizing diagnoses (i.e. DISC-YC ODD and ADHD) reported more problems. Children with an externalizing diagnosis also had reported significantly more internalizing problems on the BPI ($M = 61.4$, $SE = 0.97$) than children without such a diagnosis ($M = 57.0$, $SE = 0.77$; $F(1, 416) = 12.1$, $p < .001$, partial $\eta^2 = .03$). In contrast, children with an internalizing diagnosis did not differ

from controls in levels of BPI *internalizing problem scores*. Yet, specific internalizing diagnoses were associated with more child reported internalizing problems: children with a generalized anxiety disorder or specific phobia reported more internalizing problems. Interestingly, children with an ODD or ADHD diagnosis also had higher BPI internalizing problem scores. All effect sizes were small (Cohen, 1988), except for the moderate effect size for the difference in externalizing problems between children with and without ODD.

Interactions between child gender and internalizing or externalizing diagnoses were not significant.

In additional analyses we determined the optimal BPI symptom thresholds predicting externalizing and internalizing DISC-YC diagnoses with ROC analyses (Table S1). The threshold for an externalizing diagnosis was a BPI externalizing score above 55.7, and for an internalizing diagnosis a score above 75.0 with the respective *k* values of .25 and .14. These rather low *k* values indicated that there was little agreement between the BPI cut-offs and DISC-YC diagnoses. The cut-offs yielded a low sensitivity and somewhat higher specificity of .43 and .81, respectively for externalizing and of .21 and .91, respectively, for internalizing problems.

Children with a score in the clinical range on any of the CBCL scales, generally reported more externalizing and internalizing problems on the BPI than controls (Table 2). These associations were very consistent and also found across internalizing and externalizing domains: i.e. parent reported CBCL externalizing problem scores in the clinical range were associated with internalizing BPI problems and vice versa. We observed only two exceptions to this pattern: children with parent reported oppositional and defiant problems in the clinical range did not report more internalizing problems, and children with parent reported anxiety problems scores in the clinical range did not report more externalizing problems. Associations between parent reported and child reported externalizing problems, as well as associations between parent reported anxiety and child reported internalizing problems were of medium strength. All other effects were of small or medium-small size.

If children had reported more externalizing problems on the BPI, this was significantly associated with parent reported treatment referral ($OR = 1.05$, $p < .001$). However, when parental diagnostic information obtained with the DISC-YC was additionally added to the analyses, the association of BPI externalizing problems with treatment referral was attenuated. BPI internalizing problems were not associated with parent reported treatment referral ($OR = 1.03$, $p = .103$)

Table 1. DISC-YC diagnoses and BPI Internalizing and Externalizing scale scores

DISC-YC Diagnosis		BPI externalizing scale scores				BPI internalizing scale scores			
	n	*M (SE)*	*F (df)*	*p-value*	*partial* η^2	*M (SE)*	*F (df)*	*p-value*	*partial* η^2
Externalizing									
No	258	**50.77 (0.68)**	**19.39 (1, 416)**	**<.001**	.045	**57.04 (0.77)**	**12.08 (1, 416)**	**.001**	.028
Yes	162	**55.62 (0.86)**				**61.37 (0.97)**			
Internalizing									
No	268	50.91 (0.56)	1.29 (1, 312)	.257	.004	57.21 (0.72)	3.75 (1, 312)	.054	.012
Yes	48	53.56 (1.34)				60.82 (1.72)			
ODD									
No	265	**50.75 (0.66)**	**23.05 (1, 353)**	**<.001**	.061	**57.09 (0.75)**	**1.30 (1, 353)**	**.001**	.031
Yes	92	**57.01 (1.12)**				**62.07 (1.27)**			
ADHD									
No	265	**50.92 (0.61)**	**9.19 (1, 354)**	**.003**	.025	**57.09 (0.73)**	**10.86 (1, 354)**	**.001**	.030
Yes	93	**54.58 (1.03)**				**61.86 (1.24)**			
Mood Disorder									
No	273	50.91 (0.56)	0.38 (1, 277)	.537	.001	57.28 (0.71)	1.06 (1, 281)	.303	.004
Yes	8	52.96 (3.27)				61.63 (4.16)			
Any Anxiety Disorder									
No	269	50.88 (0.56)	1.29 (1, 309)	.256	.004	**57.18 (0.72)**	**5.37 (1, 309)**	**.021**	.017
Yes	44	52.59 (1.39)				**61.66 (1.79)**			
Specific Phobia									
No	269	50.81 (0.55)	0.003 (1, 286)	.953	.000	**57.17 (0.71)**	**7.81 (1, 286)**	**.006**	.027
Yes	21	50.69 (1.97)				**64.59 (2.56)**			

Note. Estimates are means plus *SE*s derived from ANCOVAs. Higher scores on the BPI indicate more problems. Adjusted for: child gender and child age at BPI assessment. Bold typeface emphasizes significant differences in BPI sores between participants with or without DISC-YC diagnosis.

DISCUSSION

Results of this study demonstrated that five to eight year old children from the general population with a diagnosis obtained by a structured parental interview, reported higher levels of internalizing and externalizing problems on a semi-structured child interview than children without such a diagnosis. Likewise, children with CBCL problems in the clinical range (parent report), also reported more internalizing and externalizing problems on the child interview. Further, early elementary school children's reports obtained with the BPI predict treatment referral, but not independently from DISC-YC DSM-based diagnoses obtained from parents.

Children with a DISC-YC diagnosis generally reported more problem behavior on the BPI. This is in line with results from a smaller study, indicating that mother reported DISC-YC diagnoses of anxiety and mood disorders were associated with key aspects of child reported depression and anxiety (Luby et al., 2007). We found, however, no sensitive cut-offs on the BPI to screen for DISC-YC diagnoses.

Overall, stronger associations were found between any DISC-YC diagnosis and child report of externalizing problems, than between any DISC-YC diagnosis and child report of internalizing problems.

Several explanations can be given as to why children's information on externalizing problems differed more between cases and controls than children's information on internalizing problems. First, a general challenge in obtaining reliable and valid information from young children is their developmental level. Issues like short attention span, limited vocabulary and difficulties to report on complex constructs may impact young children's reports (Measelle, Ablow, Cowan, & Cowan, 1998). This may be specifically true for the assessment of internalizing symptoms because, as argued by Luby et al. (2007), young children may have troubles to report accurately on abstract or complex constructs such as loneliness, fatigue, and worry. Second, children with externalizing problems may be more aware of their problems. Studies suggested that children who are referred for treatment become more acquainted with talking about their problems, and consequently are better able to identify their feelings (Ablow et al., 1999; Richters, 1992). Since externalizing problems are generally more disruptive than internalizing problems, and children are likely more often confronted with remarks about their behavior, they may be more conscious of their problems.

Third, a power issue may explain why children's information on externalizing problems were more strongly related with DISC-YC diagnoses than children's information on internalizing problems: there were fewer children with any of the internalizing diagnoses than with any of the externalizing diagnoses. Luby and colleagues (2003) suggested that minor changes should be made when applying DSM-IV depression criteria in an interview with preschool children, particularly changes regarding the required duration of problems.

Interestingly, ODD and ADHD DISC-YC diagnoses were related to child reports of both externalizing and internalizing problems. This is in agreement with previous studies, showing that a broad range of DSM disorders, instead of one specific diagnosis, can be related to child internalizing problems (Bellina et al., 2013; Kasius, Ferdinand, Berg, & Verhulst, 1997).

Table 2. CBCL scores above the clinical cut-off and BPI Internalizing and Externalizing scale scores

CBCL within clinical range		BPI externalizing scale scores				BPI internalizing scale scores			
	n	*M (SE)*	*F (df)*	*p-value*	*partial* η^2	*M (SE)*	*F (df)*	*p-value*	*partial* η^2
Externalizing									
No	274	**51.02 (0.67)**	**24.02 (1, 392)**	**<.001**	.058	**57.14 (0.74)**	**17.72 (1, 392)**	**<.001**	.043
Yes	122	**57.02 (1.01)**				**62.80 (1.11)**			
Internalizing									
No	272	**50.95 (0.62)**	**6.71 (1, 555)**	**.007**	.013	**57.40 (0.78)**	**7.67 (1, 555)**	**<.001**	0.41
Yes	287	**53.31 (0.61)**				**62.73 (0.76)**			
CBCL DSM ODD									
No	271	**50.86 (0.62)**	**12.11 (1, 336)**	**.001**	.035	57.25 (0.73)	2.44 (1, 336)	.119	.007
Yes	69	**55.68 (1.23)**				59.80 (1.45)			
CBCL DSM ADHD									
No	272	**50.92 (0.66)**	**17.42 (1, 413)**	**<.001**	.040	**57.14 (0.75)**	**13.74 (1, 413)**	**<.001**	.032
Yes	145	**55.66 (0.91)**				**61.90 (1.03)**			
CBCL DSM Affective									
No	270	**50.90 (0.60)**	**6.83 (1, 459)**	**.003**	.019	**57.35 (0.74)**	**11.19 (1, 459)**	**<.001**	.032
Yes	193	**53.70 (0.71)**				**61.82 (0.88)**			
CBCL DSM Anxiety									
No	274	50.92 (0.59)	1.35 (1, 368)	.197	.005	**57.30 (0.72)**	**21.04 (1, 368)**	**<.001**	.054
Yes	98	52.42 (0.99)				**63.83 (1.22)**			

Note. Estimates are means plus *SE*s derived from ANCOVAs. Higher scores on the BPI indicate more problems. Adjusted for: child gender and child age at BPI assessment. Bold typeface emphasizes significant differences in BPI sores between participants with CBCL scores within or below the clinical range

Children who scored in the clinical range on parent reported CBCL scales generally reported more externalizing and internalizing problems on the BPI. The difference between findings for externalizing and internalizing problems was not as pronounced as for DISC-YC diagnoses. One reason may be that the time interval between BPI and CBCL assessment was smaller than between BPI and DISC-YC assessment. Although studies have found fairly high stability of child reported information across time, test-retest coefficients in young children are usually not as high as for adolescent or adult self-reports (Measelle et al., 1998; Stone et al., 2013). Another reason may be that the power to find associations was larger in analyses with parent reported CBCL scales than in analyses with DISC-YC diagnoses: more children scored in the clinical range on parent reported CBCL internalizing problems than received an internalizing DISC-YC diagnosis. However, observed effect sizes, which are independent of sample size, were also stronger for associations between CBCL and BPI than for associations between DISC-YC and BPI problem scores.

Child internalizing and externalizing problems reported during the child interview was associated with treatment referral, but not independently from parental report on the DISC-YC. This suggests that the information obtained with the BPI provides no unique information on the likelihood of future treatment referral. Alternatively, treatment referral, particularly at young ages, may be highly determined by the distress or disease burden experienced by caregivers (Phares & Danforth, 1994). In particular, the BPI externalizing problems was associated with parent reported treatment referral, unless controlled for a DISC-YC externalizing diagnosis. Internalizing problems are in general less disruptive for family life, they may therefore less often lead to treatment referral. Previous studies have also reported that parents and children often disagree on the presence or absence of diagnoses based on structured diagnostic interviews (Ferdinand, 2008).

Limitations

Several limitations of the present study must be discussed. First, the interval between administration of the BPI and the structured DISC-YC interview can affect the correspondence between the measures. However, studies have shown that child reports using the BPI have a fairly acceptable stability (Measelle et al., 1998; Stone et al., 2013). Second, results of this study cannot easily be generalized to the broader population as, by the way our sample was selected, high-risk children were overrepresented. Finally, even though we used an age appropriate measure to obtain children's information of problem behavior, internal consistencies of the BPI scales were moderate as it is difficult for young children to report consistently on distinct internalizing and externalizing problems. Yet, all informant's reports are imperfect measures of children's behavior (Arseneault et al., 2005), therefore combining information from multiple sources is considered optimal (Kraemer et al., 2003).

Conclusion

The present study showed that young children's information on internalizing and externalizing problems, obtained during a structured child interview, were consistent with diagnostic information from a parent interview and with parent reported problems in a clinical range.

Children's information about problem behavior was related to externalizing and also –though to a lesser extent– to internalizing diagnostic criteria. Reports of young elementary school children may be useful as a dimensional measure to better understand problems that are relevant for a young child. However, young children's information of problem behavior obtained with the BPI has limited value in predicting later treatment referral, over and above other instruments.

REFERENCES

Ablow, J. C., & Measelle, J. R. (2003). *Manual for the Berkeley Puppet Interview: Symptomatology, Social, and Academic Modules (BPI 1.0)*. MacArthur foundation Research Network Psychopathology and Development (David J. Kupfer, Chair): The MacArthur Working Group on Outcome Assessment

Ablow, J. C., Measelle, J. R., Kraemer, H. C., Harrington, R., Luby, J., Smider, N., . . . Kupfer, D. J. (1999). The MacArthur Three-City Outcome Study: Evaluating Multi-Informant Measures of Young Children's Symptomatology. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(12), 1580-1590.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington: University of Vermont, Research Center for Children, Youth & Families.

Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffitt, T. (2005). Psychometric Evaluation of 5- and 7-Year-Old Children's Self-Reports of Conduct Problems. *Journal of Abnormal Child Psychology, 33*(5), 537-550.

Bellina, M., Brambilla, P., Garzitto, M., Negri, G. A. L., Molteni, M., & Nobile, M. (2013). The ability of CBCL DSM-oriented scales to predict DSM-IV diagnoses in a referred sample of children and adolescents. *European Child & Adolescent Psychiatry, 22*(4), 235-246.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Psychology Press.

Ferdinand, R. F. (2008). Validity of the CBCL/YSR DSM-IV scales Anxiety Problems and Affective Problems. *Journal of Anxiety Disorders, 22*(1), 126-134.

Fisher, P., & Lucas, C. (2006). *Diagnostic Interview Schedule for Children (DISC-IV) - Young Child. New York*. NY: Columbia University.

Jaddoe, V. W., van Duijn, C. M., Franco, O. H., van der Heijden, A. J., van IJzendoorn, M. H., de Jongste, J. C., . . . Hofman, A. (2012). The Generation R Study: design and cohort update 212. *European Journal of Epidemiology, 27*(9), 739-756.

Kasius, M. C., Ferdinand, R. F., Berg, H., & Verhulst, F. C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology & Psychiatry, 38*(6), 625-632.

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/Behavioral Problems in Clinic and Nonclinic Children: Correspondence Among Child, Parent and Teacher Reports. *Journal of Child Psychology & Psychiatry, 34*(6), 991-1006.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*(9), 1566-1577.

Lavigne, J. V., Lebailly, S. A., Hopkins, J., Gouze, K. R., & Binns, H. J. (2009). The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *Journal of Clinical Child & Adolescent Psychology, 38*(3), 315-328.

Luby, J. L., Belden, A., Sullivan, J., & Spitznagel, E. (2007). Preschoolers' Contribution to their Diagnosis of Depression and Anxiety: Uses and Limitations of Young Child Self-Report of Symptoms. *Child Psychiatry and Human Development, 38*(4), 321-338.

Luby, J. L., Mrakotsky, C., Heffelfinger, A., Brown, K., Hessler, M., & Spitznagel, E. (2003). Modification of DSM-IV criteria for depressed preschool children. *American Journal of Psychiatry, 160*(6), 1169-1172.

Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing Young Children's Views of Their Academic, Social, and Emotional Lives: An Evaluation of the Self-Perception Scales of the Berkeley Puppet Interview. *Child Development, 69*(6), 1556-1576.

Phares, V., & Danforth, J. (1994). Adolescents', parents', and teachers' distress over adolescents' behavior. *Journal of Abnormal Child Psychology, 22*(6), 721-732.

Rettew, D. C., Lynch, A. D., Achenbach, T. M., Dumenci, L., & Ivanova, M. Y. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*(3), 169-184.

Richters, J. E. (1992). Depressed mothers as informants about their children: a critical review of the evidence for distortion. *Psychological Bulletin, 112*(3), 485-499.

Ringoot, A. P., Jansen, P. W., Steenweg-de Graaff, J., Measelle, J. R., Ende, v. d., J., Raat, H., . . . Tiemeier, H. (2013). Young Children's Self-Reported Emotional, Behavioral, and Peer Problems: The Berkeley Puppet Interview. *Psychological Assessment, 25*(4), 1273-1285.

Stone, L. L., van Daal, C., van der Maten, M., Engels, R. C. M. E., Janssens, J. M. A. M., & Otten, R. (2013). The Berkeley Puppet Interview: A Screening Instrument for Measuring Psychopathology in Young Children. *Child & Youth Care Forum*.

Tick, N. T., van der Ende, J., Koot, H. M., & Verhulst, F. C. (2007). 14-year changes in emotional and behavioral problems of very young Dutch children. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*(10), 1333-1340.

Valla, J. P., Bergeron, L., Bérubé, H., Gaudet, N., & St-Georges, M. (1994). A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6–11 years): development, validity, and reliability. *Journal of Abnormal Child Psychology, 22*(4), 403-423.

Valla, J. P., Bergeron, L., & Smolla, N. (2000). The Dominic-R: a pictorial interview for 6- to 11-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(1), 85-93.

Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8 years). *Prevention Science, 2*(3), 165-192.

Wu, P., Hoven, C. W., Bird, H. R., Moore, R. E., Cohen, P., Alegria, M., . . . Roper, M. T. (1999). Depressive and disruptive disorders and mental health service utilization in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(9), 1081-1090; discussion 1090-1082.

Yeh, M., & Weisz, J. R. (2001). Why are we here at the clinic? Parent–child (dis) agreement on referral problems at outpatient treatment entry. *Journal of Consulting and Clinical Psychology, 69*(6), 1018.

Yesavage, J. (04-10-2002). From http://www.stanford.edu/~yesavage/ROC.html

SUPPLEMENTARY TABLE

Table S1. Cutoffs for BPI internalizing and externalizing scales derived from ROC analysis based on DISC-YC diagnoses

	Externalizing (cutoff = 55.65)		Internalizing (cutoff = 75.00)	
	High symptoms	Low symptoms	High symptoms	Low symptoms
DISC-IV-YC Internalizing/Externalizing				
Yes	69	93	10	38
No	48	210	23	245
Kappa	.25		.14	
Sensitivity	.43		.21	
Specificity	.81		.91	

chapter 4

Why mothers and young children agree or disagree in their reports of the child's problem behavior

Ringoot, A.P., Ende, van der, J., Jansen, P.W., Measelle, J.R., Basten, M., So, P., Jaddoe, V.W.V., Hofman, A., Verhulst, F.C., & Tiemeier, H. (2015)

Child Psychiatry and Human Development. Advance online publication.

ABSTRACT

This study examined multiple determinants of discrepancies between mother and child reports of problem behavior. In 5,414 6-year-olds, child problem behavior was assessed by self-report using the Berkeley Puppet Interview and by maternal report using the Child Behavior Checklist. Patterns in mother-child reports were modeled using latent profile analysis. Four profiles, differing in problem level, and the direction and magnitude of mother-child discrepancies, were identified: one profile representing agreement (46%), another representing slight discrepancies (30%), and two representing higher problem levels and more discrepancies. In the latter two profiles either children (11%) or mothers (13%) reported more problems. Compared to the first profile, the second was predominantly characterized by a positive family environment, the third by child cognitive difficulties, and the fourth by harsh discipline and poor family functioning. Knowledge about specific child or family characteristics that contribute to mother-child discrepancies can help to interpret informants' reports and to make diagnostic decisions.

INTRODUCTION

The assessment of child emotional and behavioral problems often relies on the use of multiple informants. Information from several sources, such as parents, teachers, clinicians and the children themselves, provides a more accurate description of children's problems (Kraemer et al., 2003). Recently, instruments have become available to obtain structured self-reported information from children as young as four years old (Ablow et al., 1999; Ringoot et al., 2013; Valla, Bergeron, & Smolla, 2000). Hence, young children's information can be included in studies that usually relied on adult information only, and multi-informant approaches can be applied from an early age onwards.

Although multi-informant approaches are a common practice, information obtained from multiple sources is often discrepant (Achenbach, McConaughy, & Howell, 1987; Youngstrom, Loeber, & Stouthamer-Loeber, 2000). A meta-analysis from Achenbach and colleagues (1987) and more recent studies demonstrated that there is only little agreement across informants' reports of problem behavior (De Los Reyes & Kazdin, 2004; Grills & Ollendick, 2002; Youngstrom et al., 2000). In addition, discrepancies between informants' reports are found across the lifespan (van der Ende, Verhulst, & Tiemeier, 2012). These discrepancies can influence how researchers and clinicians interpret information and may thus impact decision making in research and clinical practice (De Los Reyes & Kazdin, 2004; De Los Reyes, Thomas, Goodman, & Kundey, 2013). Consequently, discrepancies in informant reports have often been considered a nuisance and a problem of measurement error (De Los Reyes et al., 2013). Yet, there can be useful information in the diverging reports of informants. Research has therefore aimed at identifying determinants of these discrepancies, as this may help to better value information of each individual informant.

Discrepancies can reflect the setting or context where information was obtained (e.g. home vs. non-home), as contextual fluctuations in the expression of children's behavior may result in differing reports. In addition, discrepancies can reflect the underlying perspectives of informants about what is normal and abnormal behavior (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005; Verhulst & van der Ende, 2008). While both context and perspective may account for discrepancies, a recent review on the use of multi-informant reports encouraged researchers to focus on factors other than context when examining informant discrepancies (De Los Reyes et al., 2013). Further, parents and young children are the focus of the present study, because parents (in particular mothers) are the most prominent informants about young children's problem behavior. Parents and young children likely report largely about similar contexts.

Over the past decades research has identified several characteristics that may impact parent's and children's perspectives and that are associated with parent-child discrepancies. Among these are demographic and socioeconomic characteristics of families like children's age, gender and family socioeconomic status (Achenbach et al., 1987; Chi & Hinshaw, 2002; Duhig, Renk, Epstein, & Phares, 2000; Grills & Ollendick, 2003; Kolko & Kazdin, 1993); parental psychopathology, ranging from parental depression and anxiety to anger and hostility (Boyle & Pickles, 1997; Briggs-Gowan, Carter, & Schwab-Stone, 1996; Chi & Hinshaw, 2002; Frick, Silverthorn, & Evans, 1994; Krain & Kendall, 2000; Richters, 1992; Treutler & Epkins, 2003; Youngstrom, Izard, & Ackerman, 1999); and

dysfunctional interactions between family members (De Los Reyes & Kazdin, 2006). Specifically, problems in family functioning, parent-child conflict, or harsh discipline may influence family members' perceptions of problem behavior (De Los Reyes, Youngstrom, Pabon, Youngstrom, Feeny et al., 2011; Goodman, De Los Reyes, & Bradshaw, 2010; Grills & Ollendick, 2003; Kolko & Kazdin, 1993; Reid, Kavanagh, & Baldwin, 1987; Youngstrom et al., 2000). In contrast, child cognitive abilities have rarely been studied as determinants of informant discrepancies (Penney & Skilling, 2012). Differences in cognitive functioning may, however, determine the developing child's ability to report on emotions and behavior.

A substantial number of characteristics have thus been studied in association with parent-child discrepancies, yet results remain inconclusive (De Los Reyes & Kazdin, 2005) and it is unclear which factors independently determine informant discrepancies. One reason for this gap is that most studies focused on single determinants of parent-child discrepancies, while only few tested the independent contribution of multiple determinants (e.g. De Los Reyes & Kazdin, 2006; De Los Reyes et al., 2011; Treutler & Epkins, 2003; Youngstrom et al., 2000). Further, studies that did test multiple determinants of parent-child discrepancies were generally conducted in relatively small samples or in samples of clinically referred children and their parents. Another gap in the existing literature is that most studies were performed in children from age eight onwards. A number of recent studies did examine informant discrepancies in younger children (Berg-Nielsen, Solheim, Belsky, & Wichstrom, 2012; Grietens et al., 2004; Lavigne, Dahl, Gouze, LeBailly, & Hopkins, 2014; Strickland, Hopkins, & Keenan, 2012). However, given the paucity of methods to obtain young children's self-reports, these studies focused on parent-teacher discrepancies. The studies showed that interrater agreement is also low when examining young (i.e. preschool) children's problem behavior (Berg-Nielsen et al., 2012; Grietens et al., 2004; Lavigne et al., 2014). One study found that parent-teacher agreement appeared stronger when parents and teachers reported about the same context (e.g. school) (Strickland, Hopkins, & Keenan, 2012). In addition, family and parenting characteristics, such as parental stress level and family conflict, determined that parents reported more problems than teachers (Berg-Nielsen et al., 2012; Lavigne et al., 2014). Yet, none of these studies on informant discrepancies included the child's perspective of their problems. Kraemer et al. (2003) were the first to include young children's self-reports in a study related to the issue of informant discrepancies, but they used children's information to illustrate a method on combining multi-informant reports. To the best of our knowledge, no study examined independent determinants explaining parent-child discrepancies in children younger than eight years old.

The present study on informant discrepancies expands the age range of previous studies by focusing on children aged five to seven years. Further, we examined the unique contribution of a wide range of determinants that may contribute to discrepancies between mother report and young child self-report.

Few validated instruments are available to systematically assess behavioral and emotional problems in children younger than eight years, thus we could not obtain maternal information using different versions of the same instrument. Consequently, we examined patterns in the relative problem levels (z-scores) that mothers and children reported. Reporting patterns were obtained

with latent profile analysis (LPA), which is a way to explore clusters of symptoms across informants. Previous studies have also used and recommended latent variable approaches in multi-informant studies (De Los Reyes et al., 2013; De Los Reyes et al., 2011; Goodman et al., 2010; Rubin, Althoff, Walkup, & Hudziak, 2013). Our approach allowed us to examine the *magnitude of discrepancies* between a child's and a mother's relative problem level, the *direction of discrepancies* –i.e. whether the child or the mother reported more problems– (De Los Reyes et al., 2011), and the *relative level of problems* that mother and child reported. After obtaining these mother-child reporting profiles, we studied determinants of these profiles. Specific hypotheses regarding the associations could not be formed, given inconsistent results of prior research and the lack of studies examining determinants of discrepancies using young child self-reports. However, in general, we hypothesized that profiles with lager discrepancies and higher problem levels would be determined by problems in cognitive functioning of children, and by adverse socio-demographic and family related characteristics.

METHODS

Design and Study Sample

This study was embedded in the Generation R Study, a population-based prospective cohort investigating growth, development and health from fetal life onward. The design and data collection procedures have previously been described in detail (Jaddoe et al., 2012). All children were born between April 2002 and January 2006 in Rotterdam, the Netherlands. Typically, enrolment took place in early pregnancy. The study was approved by the Medical Ethics Committee of Erasmus Medical Center, Rotterdam. Written informed consent was obtained from all caregivers.

Participants with consent for the study-phase from age 4 to 16 were eligible for the present study. Consent for this particular study phase was obtained from caregivers of 8,305 children that had all been followed since birth. Of the 8,305 children eligible for follow-up examinations, 6,690 children aged 5 to 7 years old ($M = 6.1$, $SD = 0.04$) visited, together with their caregiver, our research center at the Erasmus Medical Center–Sophia Children's Hospital. During this visit child self-report of emotional and behavioral problems was obtained using the Berkeley Puppet Interview (BPI; n = 6,598). We excluded BPI data if coders had indicated that the child did not understand the interview, or if >25% of the data was missing on all scales (n = 77). We also excluded data if only BPI and no parent reported Child Behavior Checklist (CBCL) data was available (n = 1,022) and if children were eight years or older when either the BPI or the CBCL was completed (n = 85). This resulted in a sample for analysis of 5,414 (65%) children.

Characteristics of participants that were included in the study (responders; n = 5,414), were compared with characteristic of participants with missing data on child or mother report (non-responders; n = 2,891). Among the non-responders, the average maternal age was lower ($M1$ = 28.80 and $M2$ = 31.07, t = -18.47 (5,254), $p < .001$). Non-responding mothers were more often of non-Western origin (51.8% vs. 32.2%, $\chi^2(1) = 280.6$, $p < .001$) and lower educated (66.3% vs. 43.4%, $\chi^2(1) = 356.2$, $p < .001$). No differences were detected in general maternal psychopathology scores

(*M1* = 3.70, *M2* = 3.35, *t* = 1.59 (1,310), *p* = .113) or family income (26.1% vs. 23.1%, $\chi^2(1) = 3.6$, *p* = .059) between non-responders and responders.

Measures

Berkeley Puppet Interview – Child reported problem behavior was assessed with the Berkeley Puppet Interview, a semi-structured interactive interview to obtain self-reports from young children (Ablow & Measelle, 2003). During the interview at the research center, two identical dog hand puppets, 'Iggy' and 'Ziggy', were introduced to the child and invited him or her to engage in conversation. The puppets made opposing statements about themselves and asked children to indicate which statement described him or her best. Afterwards, videotaped interviews were scored by trained coders on a 7-point scale ranging from 1 to 7. The exact score depended on which of the puppets' statements the child chose and how much emphasis was put on the answer. Higher scores represented more problems. Items were summed to compute scale scores. We used the internalizing and externalizing scales. The internalizing scale (20 items) comprised the Depression, Separation Anxiety and Overanxious scales. The externalizing scale (21 items) comprised the Oppositional Defiant, Overt Hostility and Conduct Problems scales. A recent psychometric study of the BPI in the Generation R Study suggested that the interview has a multidimensional structure, acceptable internal consistencies, and that scores could be interpreted validly as shown by associations with socio-demographic factors (Ringoot et al., 2013).

Child Behavior Checklist 1½-5 – Parent reported child emotional and behavioral problems were assessed with the Child Behavior Checklist for ages 1½-5 years (CBCL; Achenbach & Rescorla, 2000). The primary caregiver completed the questionnaire at home, prior to the research center visit. As the majority (60%) of the children were younger than 6 years old at the time of CBCL assessment (37% were 6 years old and 3% of the children were 7 years old), we used the CBCL 1½-5 version for all children during this assessment wave to enhance comparability across all children, as recommended in the ASEBA manual (Achenbach & Rescorla, 2000). The primary caregiver rated the occurrence of their child's behavior within the past two months on a scale from 0 (not true) to 2 (very true or often true). In 92% of the sample the primary caregiver was the mother, therefore we refer to this report with the term 'mother report'. A sensitivity analysis was performed using mother reported data only. Items were summed to compute scale scores, with higher scores representing more problems. We used the 36-item internalizing and 24-item externalizing scales. Good reliability and validity have been reported for the CBCL (Achenbach & Rescorla, 2000). Internal consistencies of the internalizing and externalizing scales in the present study were α = .86 and α = .90, respectively. These internal consistencies were similar in children younger than 6 years and in children older than 6 years.

Determinants of informant discrepancies – The choice of potential determinants of discrepancies was based primarily on prior studies (De Los Reyes & Kazdin, 2005, 2006; De Los Reyes et al., 2011; Duhig et al., 2000). *Demographic and socioeconomic characteristics* (maternal age at pregnancy,

child age, child gender, siblings, maternal national origin, maternal education, marital status, and family income) were obtained from questionnaire data and medical records. Maternal age and child gender were obtained from medical records completed by midwives and gynecologists. Presence of siblings in the household, maternal national origin, maternal education, and family income were obtained by questionnaires filled out prior to the research center visit at the same time as the CBCL assessment. Maternal national origin was classified by the countries of birth of her parents, according to the Dutch standard classification criteria (Statistics Netherlands, 2004), and was categorized into "Dutch", "other-Western" (i.e. other European, North-American and Oceania) and "non-Western" (i.e. Turkish, Moroccan, Indonesian, Cape Verdean, Surinamese and Antillean) national origin. Maternal education was defined as highest completed schooling and classified in three levels; "primary", "secondary" or "higher" education. The mothers' marital status was categorized as "married or cohabiting" and "single parenthood". Family income, defined by the total net month income of the household, was categorized as "<1200€ (social security level)", "1200-2000€ (below modal)", "2000-4000€ (above modal)" and ">4000€ (two times above modal)" net income per month.

Parental psychopathology (depression, anxiety and hostility) was assessed using the Dutch version of the Brief Symptom Inventory (BSI); a validated self-report questionnaire which is widely used in clinical and research settings (Derogatis, 1993). The BSI was completed by each parent when the child was 3 years old. The BSI originally consists of 53 items, in the present study we used the depression (6 items), anxiety (6 items) and hostility (5 items) subscales. Mothers and fathers rated the extent to which each item described their feelings in the past week according to a 5-point scale ranging from 0 (not at all) to 4 (extremely). We calculated a mean score on the subscales, with higher scores representing more problems (Cronbach's alphas for mothers and fathers, respectively: depression α = .83 and α = .79; anxiety α = .75 and α = .63; hostility α = .62 and α = .60). *Family and parenting factors* that were considered were family functioning and harsh discipline. *Family functioning* as perceived by the primary caregiver (in 92% this was the mother) was measured at child age 6 years with the General Functioning scale of the McMasters Family Assessment Device (FAD: Miller, Epstein, Bishop, & Keitner, 1985), a validated 12-item self-report measure of family health and pathology. The primary caregiver rated how well each item described their family on a 4-point scale ranging from 1 (disagree strongly) to 4 (agree strongly). Sample items include "We cannot get along well with each other" and "There are a great many unpleasant, painful feelings in the family". A mean score on the 12 items was calculated, with higher scores indicating more problems in family functioning (α = .89). *Disciplinary styles* of mothers and fathers were assessed using an adapted version of the Parent–Child Conflict Tactics Scale (Straus et al., 1998), when children were three years old. Parents rated their use of discipline during the past two weeks on a 6-point scale ranging from "never" to "5 times or more". The categories "twice", "3 times", "4 times", and "5 times" were combined because of low prevalence rates. This resulted in a 3-point scale from 0 (never) to 2 (twice or more). In the Generation R Study, factor analysis of this adapted CTS-PC has identified a harsh discipline construct consisting of 6 items (Jansen et al., 2012). Items were included like "I shouted or screamed angrily at my child" and "I angrily pinched my child's arm". We

calculated a harsh discipline sum score by summing the scores on the 6 'harsh discipline' items (α = .57 for mothers and α = .55 for fathers). These modest Cronbach's alphas reflect the small number of items in the scale, that in addition target a range of parental behaviors that do not necessarily co-occur (Cortina, 1993).

Information on *children's cognitive abilities* (IQ and verbal ability) was obtained during the research center visit. *IQ* was assessed using two subtests of a Dutch non-verbal IQ test: Snijders-Oomen Niet-verbale Intelligentie Test–Revisie (SON-R 2½-7: Tellegen, Winkel, Wijnberg-Williams, & Laros, 2005). The two subsets used were 'Mosaics', which taps into spatial visualization abilities, and 'Categories', which assesses abstract reasoning abilities. The raw test scores were converted into non-verbal IQ using norms that were tailored to exact age (Tellegen et al., 2005). The correlation of the IQ score derived by the mosaics and categories subsets and the IQ scores derived by the total test was high, r = .86 in an unrelated sample of 626 children aged 4½-7½ years (Tellegen, personal communication). To investigate language development in children, we assessed *verbal ability* with a receptive language subtest (syntax choice task) of a Dutch test battery (Van Bon & Hoekstra, 1982). A sum score on this subtest was calculated, with higher scores indicating better performance (α = .67).

Statistical analysis

Baseline characteristics of the data were explored and Pearson's correlation coefficients between BPI and CBCL internalizing and externalizing scale scores were calculated. Raw scores on the BPI and CBCL scales were transformed with inverse and square root transformations respectively, to achieve a more normal distribution of the data. The skewness of BPI and CBCL after the transformation was -0.16 and 0.23 for the internalizing scales, and 0.53 and -0.04 for the externalizing scales.

Next, *z*-scores of these transformed scale scores were computed. *Z*-scores have previously been used in multi-informant studies to compute a discrepancy measure, namely: 'standardized difference scores' (SDS': subtracting the z-standardized scores of two informants; De Los Reyes & Kazdin, 2004). While SDS' provide a direct way to study discrepancies, results can be difficult to interpret. One single score represents the ratings from two different informants, thus information on the initial problem level that the informants reported is lost. Also, different problem levels can result in the same SDS. For example, if mother and child agreed that there were relatively many problems or if they agreed that there were none, this could both result in a SDS of 0, while these same SDS' may well have different determinants. In the present study, we therefore used actual z-scores instead of a unidimensional discrepancy measure like SDS'.

To study determinants of informant discrepancies, data were analyzed in two steps. In the first step, patterns in mother and child report of problem behavior were modeled using latent profile analyses (LPA). *Z*-scores of the transformed internalizing and externalizing BPI and CBCL scale scores were used as indicators. LPA is a 'person-centered' method using continuous variables as indicators. The method attempts to identify profiles of individuals that show similar patterns across the indicators. A maximum likelihood estimator with robust standard errors was chosen to take into account any remaining skewness in the data. We started with a one profile model and increased the

number of profiles until a parsimonious model with good model fit was reached. The best fitting model was determined by the sample size adjusted Bayesian Information Criterion (BIC), Akaikes Information Criterion (AIC), entropy, Vuong-Lo-Mendell-Rubin Likelihood Test (VLMRT) and the Bootstrapped Likelihood-Ratio Test (BLRT; Nylund, Asparouhov, & Muthén, 2007). Other relevant information such as interpretability and latent group size was also taken into account. Once the number of latent profiles was determined, each participant was assigned to the profile for which they had the highest assignment probability. LPA was conducted with Mplus version 6.2 (Muthén & Muthén, 1998–2010).

In the second step, associations between potential determinants of discrepancies and the latent profiles were studied using multinomial regression analyses. We first examined univariate associations of demographic and socioeconomic characteristics, caregiver psychopathology, family and parenting factors, and cognitive abilities of the children with profile membership. Next, to study the independent contribution of each determinant, we added all determinants to a multinomial regression model. To facilitate the interpretation of the findings and to enhance comparability between the variables, the continuous measures of BSI, FAD, harsh discipline and language development were divided by their standard deviation (*SD*-scores). All regression analyses were performed in SPSS, version 20.0 (IBM Corp, Armonk, NY).

Missing values on the potential determinants of discrepancies were imputed using Multiple Imputation in SPSS 20.0. Missing values on these determinants ranged from < 0.01% for gender, age and ethnicity to 14% for verbal ability of the child (mean percentage missing data = 4.4%). For logistical reasons, participants living in the south of Rotterdam were not approached for participation in the preschool period (0-4 years; n = 1,166). As a result, there was more missing data on information obtained by questionnaires in that period, i.e. parental psychopathology and harsh discipline: missing data ranged from 30% for maternal harsh discipline to 43% for paternal psychopathology. Imputed data were estimated using all predictors in the model, as well as mother, child and teacher reported problem behavior, and prenatal reports of psychopathology of both parents. Regression coefficients were averaged over twenty imputed datasets. To examine whether the imputation of missing data had influenced the results, analyses were repeated in the complete cases. In addition, analyses were repeated using only mother reports of child problem behavior, instead of primary caregiver reports.

In additional analyses, we examined whether discrepancies were not merely a result of potential biases in mothers' reports, or of children's inability to accurately report their problems. To address this issue, the discrepancy profiles were related to ratings of an external informant, in this case the teacher. Teacher reports of children's problems were obtained using the validated Teacher Report Form (Achenbach & Rescorla, 2001). This information was available in a subgroup of 3,102 participants, aged 6.8 years (*SD* = 2.7). The obtained profiles were associated with ratings of the teacher, using linear regression analysis. Demographic and socioeconomic characteristics, caregiver psychological problems, family and parenting factors, and cognitive abilities of the child were added as covariates.

Shared-method variance may have inflated results because the majority of the family and parenting measures were reported by mothers. We therefore performed additional analyses to examine whether an objective measure of parenting, namely observed maternal sensitivity, was associated with the discrepancy profiles. We used a maternal sensitivity measure obtained during a home visit when the child was 4 years old. This was the most recent assessment of maternal sensitivity. At the home visit, mother and child were asked to perform two interaction tasks that were too difficult for the child: building a tower and an etch-a-sketch task. The subsample and coding of the tasks was previously described in more detail (Kok et al., 2013). Observed maternal sensitivity was available in a subgroup of 553 participants, aged 4.3 years (*SD* = 0.1) and was included as a determinant in nominal regression analyses, adjusted for demographic and socioeconomic covariates.

RESULTS

Baseline characteristics of the children and their mothers are presented in Table 1. The majority of mothers were of Dutch national origin (60%), their mean age was 31 years (*SD* = 4.9). About 44% of the mothers had a medium or low education, and 23% of the families had a household income below modal.

Pearson's correlation coefficients between child self-report on the BPI and maternal report on the CBCL were low, but highly significant; internalizing $r = .14$, $p < .001$; externalizing $r = .19$, $p < .001$ (see also Ringoot et al., 2013).

The model fit for the LPA solutions for one to five profiles is shown in Table 2. Although BIC and AIC still decreased from a four to a five profile solution, the VLMRT pointed towards four profiles as best solution. Also, this four profile solution resulted in distinct and interpretable profiles.

The four profile solution is graphically represented in Figure 1. The profiles that we identified could be distinguished by three features: 1) the relative level of problems, 2) the magnitude of discrepancies and 3) the direction of the discrepancies, indicating whether the child or the mother reported more problems. The *first* and largest profile (*n* = 2,474; 46%) was characterized by *agreement. Children and mothers both reported an average relative problem level* (both had *z*-scores of about 0), although mothers mostly reported a slightly higher problem level than their children. This profile was labeled 'Child = Mother'. The *second* profile (*n* = 1,634; 30%) was characterized by *slight discrepancies*, children represented in this profile reported an average relative problem level (*z*-scores of about 0), while mothers reported a lower relative problem level (negative *z*-scores). Overall, the lowest problem levels were reported in this second profile. We labeled this profile 'Child > Mother'. The third and fourth profiles were characterized by higher relative problem levels and more discrepancies. The direction of discrepancies in these profiles differed. Characteristic for the *third* profile (*n* = 614; 11%) was that *children reported a higher relative problem level than mothers* (child *z*-scores ≥ 1 *SD*; mother z-scores about 0). Children represented in this third profile reported particularly high levels of externalizing problems (≥ 1.5 *SD*), and lower, albeit still high, levels of internalizing problems (≈ 1 *SD*). This profile was labeled 'Child >> Mother'. Characteristic for the *fourth* profile (*n* = 692; 13%) was that *children*

reported an average relative problem level, while mothers reported a higher relative problem level (child *z*-scores about 0, mother *z*-scores > 1), this profile was labeled 'Child << Mother'.

Table 1. Baseline characteristics of original data

N = 5,414	Mean (*SD*)[a]
Parental characteristics	
Maternal age at pregnancy, years	31.07 (4.86)
National origin (mother) (%)	
Non-Western	32.2
Other-Western	8.5
Dutch	59.3
Education mother (%)	
Primary	4.1
Secondary	39.3
High	56.6
Marital status mother (% Single Parenthood)	13.6
Household income (%)	
<1200	6.9
1200-2000	16.2
2000-4000	42.8
>4000	34.1
Maternal psychopathology score (BSI)	
Depression	0.13 (0.32)
Anxiety	0.18 (0.31)
Hostility	0.18 (0.27)
Paternal psychopathology score (BSI)	
Depression	0.09 (0.25)
Anxiety	0.15 (0.25)
Hostility	0.16 (0.25)
Family Functioning score (FAD)	1.52 (0.42)
Maternal harsh discipline score	2.16 (1.91)
Paternal harsh discipline score	1.81 (1.82)
Child characteristics	
Gender (% boys)	49.8
Siblings (% no)	18.6
Age BPI (months)	73.38 (4.55)
Age CBCL (months)	72.21 (4.55)
Absolute time interval BPI-CBCL months (median, (90% range)])	1.06 (0.03-3.9)
IQ	102.10 (14.69)
Language development score	22.09 (2.97)

[a] Presented are mean and standard deviation, unless otherwise indicated.

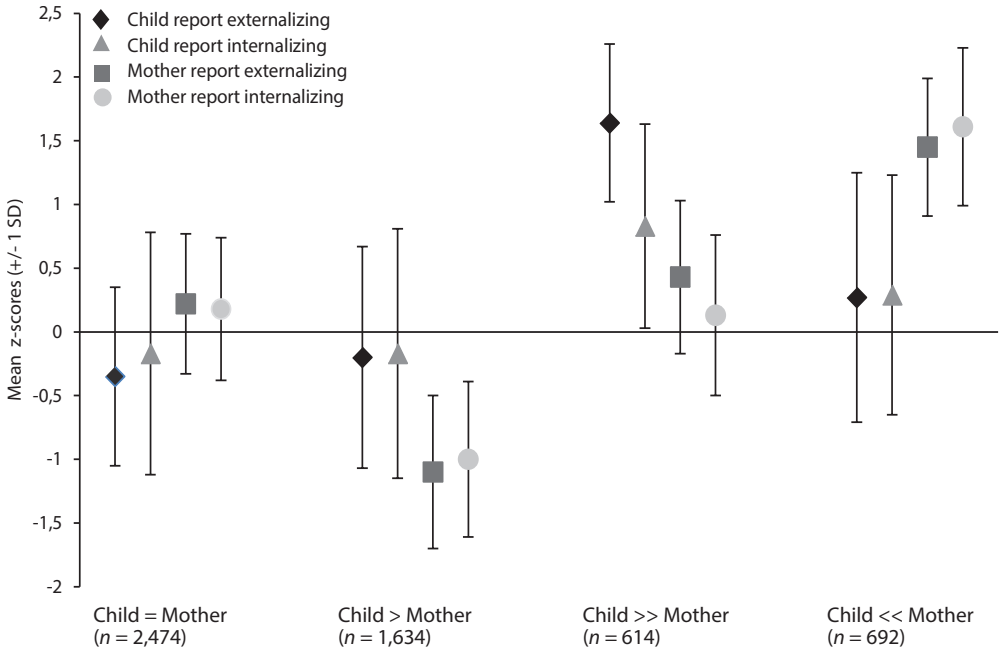
Table 2. Goodness of fit for latent profiles of child and mother reported child problem behavior

N = 5,414	BIC	AIC	Entropy	*p*-value, VLMRT	*p*-value, BLRT
1 profile	61,340.34	61,312.98	NA	NA	NA
2 profiles	58,966.82	58,922.37	.65	<.001	<.001
3 profiles	57,999.63	57,938.08	.72	<.001	<.001
4 profiles	**57,663.55**	**57,584.91**	**.69**	<.001	<.001
5 profiles	57,373.66	57,277.92	.69	.127	<.001

Note. Bold typeface indicates preferred model. BIC=Sample size adjusted Bayesian Information Criterion, AIC=Akaikes Information Criterion, VLMRT= Vuong-Lo-Mendell-Rubin Likelihood Test, BLRT=Bootstrapped Likelihood-Ratio Test.

As discussed above, we considered the 4 profile solution to fit the data best. No distinct profile was found where mothers and children both reported a high level of problems. The mother-child dyads agreeing that relative problem levels were high (both ≥ 1 *SD*), were mostly found in the third and fourth profiles: 69% of the mother-child dyads that agreed on a high level of internalizing problems were found in the fourth profile, 16% in the third profile. Also, 57% of the mother-child dyads that agreed on a high level of externalizing problems were in the fourth and 42% in the third profile.

Figure 1. Reporting patterns within each latent profile



Next, the determinants of profile membership were examined. We studied unadjusted associations of potential determinants with profile membership and also the independent contribution of each of these determinants in a multivariate multinomial regression. The profile with most members ('Child = Mother') was chosen as reference group in the multinomial regressions, this was also the profile with most agreement between child and mother. The results for the univariate multinomial regressions are presented in supplementary material (Table S1). Briefly, most determinants were associated with membership of one or more of the profiles. As compared to the reference group, more favorable determinants characterized the 'Child > Mother' profile. In contrast, more adverse determinants were associated with the 'Child >> Mother' and 'Child << Mother' profiles.

In Table 3, results from the multivariate multinomial logistic regression are shown. All profiles were compared to the reference group 'Child = Mother'. In the multivariate analysis many associations observed in the univariate analyses were attenuated and the pattern of associations for each profile became more distinct. First, we discuss determinants of the 'Child > Mother' profile. Mothers assigned to this profile were on average older at childbirth (OR = 1.02, p < .05), children were more likely to be a girl (OR_{boy} = 0.86, p < .05) and to have siblings ($OR_{no\ siblings}$ = 0.78, p < .05). Parental psychopathology was not associated with membership of this profile, as was the case in most of the other profiles. Problems in family functioning and harsh discipline were less likely in members of this profile (OR_{FAD} = 0.71, p < .001; $OR_{HD\text{-}mother}$ = 0.77, p < .001; $OR_{HD\text{-}father}$ = 0.89, p < .05). Further, cognitive abilities of children within this group were no different from those of children in the reference group. Then, we examined determinants of the two profiles with larger mother-child reporting discrepancies: the 'Child >> Mother' and 'Child << Mother' profiles. Children assigned to the 'Child >> Mother' profile were more often boys than girls (OR = 1.99, p < .001). Fathers of children assigned to this profile had reported higher levels of harsh discipline (OR = 1.15, p < .05). Most salient was that child cognitive abilities determined membership of this profile; children were more likely to have a lower IQ and poor performance on a receptive language task (OR_{IQ} = 0.98, p < .001; $OR_{language}$ = 0.81, p < .001). Finally, children assigned to the 'Child << Mother' profile were more often boys than girls (OR = 1.50; p < .001). Mothers had more often a lower level of education (OR = 1.63, p < .05) and they reported higher levels of anxiety (OR = 1.15; p < .05). In addition, problems in family functioning and harsh discipline were more likely in members of this profile (OR_{FAD} = 1.38, p < .001; $OR_{HD\text{-}mother}$ = 1.16, p < .01; $OR_{HD\text{-}father}$ = 1.15, p < .05). We also found that children in this group were more likely to have a lower IQ than children in the reference group (OR = 0.99, p < .01).

Results from the multinomial logistic regressions with imputed data were compared to results obtained with the original, non-imputed, data. Results were essentially unchanged, with the exception that there was an effect of ethnicity in the original data that was not present in the imputed data. All analyses were also repeated using mother reported child problem behavior data only, instead of primary caregiver data. Results for these analyses were largely similar, with the only differences that the effect of maternal age in the 'Child > Mother' profile and the effect of maternal anxiety in the 'Child << Mother' profile disappeared.

Table 3. Adjusted associations of potential determinants with profile membership

	Child = Mother *(1st profile)* *n* =2,474, *Ref.*	Child > Mother *(2nd profile)* *n* = 1,634			Child >> Mother *(3rd profile)* *n* = 614			Child << Mother *(4th profile)* *n* = 692		
	Mean / %[a]	*M* / %[a]	*OR*	95%*CI*	*M* / %[a]	*OR*	95%*CI*	*M* / %[a]	*OR*	95%*CI*
Demographic and Socioeconomic Characteristics										
Maternal age at pregnancy	31.1	31.6	1.02*	1.00; 1.03	30.6	1.00	0.98; 1.02	30.0	0.98	0.96; 1.00
Child age	6.09	6.07	1.00	0.99;1.02	6.1	1.00	0.98;1.02	6.1	0.99	0.97;1.01
Child Gender										
Boy	47.9	43.3	0.86*	0.75; 0.98	65.1	1.99***	1.65; 2.40	58.4	1.50***	1.25; 1.80
Girl	52.1	56.7	1		34.9	1		41.6	1	
Siblings										
No	19.4	15.1	0.78*	0.65; 0.94	21.7	1.05	0.82; 1.33	23.4	1.10	0.88; 1.39
Yes	80.6	84.9	1		78.3	1		76.6	1	
Ethnicity										
Non-Western	30.5	25.3	1.03	0.87;1.22	40.1	1.07	0.85; 1.34	47.5	1.15	0.92; 1.43
Other-Western	8.4	9.1	1.18	0.93; 1.49	8.8	1.11	0.79; 1.54	7.4	0.96	0.68; 1.35
Dutch	61.1	65.6	1		51.1	1		45.1	1	
Educational level										
Primary	3.3	3.0	1.37	0.90; 2.07	4.6	0.80	0.48; 1.33	9.5	1.63*	1.07; 2.49
Secondary	39.0	34.0	0.98	0.84; 1.16	45.2	0.95	0.76; 1.18	47.8	1.13	0.91; 1.40
High	57.7	63.0	1		50.2	1		42.6	1	
Marital Status										
Single parenthood	13.2	10.1	1.06	0.82; 1.38	17.9	1.11	0.81; 1.52	20.9	1.06	0.79; 1.42
Married/cohabiting	86.8	89.9	1		82.1	1		79.1	1	

Table 3. continued

Income										
<1200	6.6	4.3	0.82	0.55; 1.22	9.7	1.35	0.84; 2.17	13.2	1.33	0.85; 2.10
1200-2000	14.9	12.9	1.09	0.83; 1.42	21.5	1.36	0.96; 1.94	24.4	1.17	0.82; 1.65
2000-4000	44.1	41.4	0.97	0.83; 1.14	43.7	1.15	0.90; 1.47	40.4	1.02	0.79; 1.31
>4000	34.4	41.4	1		25.1	1		22.0	1	
Parental Psychopathology[b,c]										
Depression mother	0.15	0.10	1.06	0.93; 1.21	0.19	1.07	0.93; 1.24	0.27	1.00	0.88; 1.15
Anxiety mother	0.20	0.15	0.92	0.82; 1.03	0.23	1.00	0.88; 1.14	0.33	1.15*	1.02; 1.29
Hostility mother	0.21	0.15	0.91	0.80; 1.02	0.23	0.99	0.87; 1.13	0.30	1.05	0.94; 1.17
Depression father	0.13	0.10	0.98	0.88; 1.09	0.16	1.01	0.90; 1.13	0.19	1.00	0.91; 1.11
Anxiety father	0.20	0.16	0.92	0.83; 1.01	0.22	1.01	0.88; 1.16	0.26	1.08	0.97; 1.20
Hostility father	0.21	0.17	1.00	0.90; 1.12	0.24	1.01	0.87; 1.18	0.26	0.98	0.87; 1.09
Family and Parenting[b,c]										
Family functioning problems	1.5	1.4	0.71***	0.66; 0.77	1.6	1.03	0.93; 1.13	1.7	1.38***	1.26; 1.51
Harsh discipline mother	2.3	1.8	0.77***	0.70; 0.85	2.7	1.06	0.94; 1.19	3.2	1.16**	1.04; 1.30
Harsh discipline father	2.0	1.5	0.89*	0.81; 0.98	2.4	1.15*	1.01; 1.30	2.6	1.15*	1.02; 1.29
Cognitive abilities of the child[d]										
IQ	102.9	104.1	1.00	1.00; 1.01	97.4	0.98***	0.98; 0.99	98.2	0.99**	0.98; 1.00
Language[b]	22.24	22.47	1.04	0.96; 1.13	21.13	0.81***	0.73; 0.89	21.39	0.92	0.83; 1.02

[a] Presented are raw and un-adjusted scores; [b] Odds ratios are presented 'per *SD* change in the determinant'; [c] Higher values represent more problems; [d] Higher values represent fewer problems; *$p < .05$, **$p < .01$, ***$p < .001$

Additional analyses using teacher reported child problems showed that children in the 'Child > Mother' profile had less internalizing and less externalizing problems according to teachers (*B* = -0.24, *95% CI* = -0.34; -0.15, $p < .001$ and *B* = -0.13, *95% CI* = -0.24; -0.03, $p < .05$) than children in the reference profile ('Child = Mother'). Conversely, children in the 'Child << Mother' profile had more teacher reported internalizing and externalizing problems than children in the reference profile (*B* = 0.19, *95% CI* = 0.06; 0.33, $p < .01$ and *B* = 0.39, *95% CI* = 0.24; 0.53, $p < .001$) consistent with the high problem level mothers reported. According to teacher ratings, children classified in the 'Child >> Mother' profile had more externalizing problems (*B* = 0.41; *95% CI* = 0.26; 0.56, $p < .001$), but not more internalizing problems than children in the reference profile. This specific association with externalizing problems was consistent with the children's own report. Thus, if either mother or child reported a higher problem level, these problems were, on average, also apparent for teachers.

Additional analyses in a subgroup of 553 mothers and their children indicated that higher levels of observed maternal sensitivity were more common in the 'Child > Mother' profile (*n* = 184, mean sensitivity z-score = 0.19 versus mean sensitivity z-score = -0.08 in the 'Child = Mother' profile (*n* = 269), *OR* = 1.13, *95% CI* = 1.06; 1.58, $p < .01$). Observed maternal sensitivity was not associated with membership in any of the profiles (with fewer participants) representing more discrepancies, but associations were in the expected direction ('Child >> Mother', mean = -0.11, *OR* = 0.95, *95% CI* = 0.70; 1.28, *n* = 56; 'Child << Mother', mean = -0.23, *OR* = 0.82, *95% CI* = 0.58; 1.15, *n* = 43).

DISCUSSION

In this large population-based sample we examined independent determinants of discrepancies between maternal and young children's self-report of problem behavior. To this aim we first identified patterns within the problem behavior reports of mothers and their children. Child-related characteristics, and in particular cognitive difficulties, predicted that children reported more problems than their mothers. In contrast, adverse family and parenting characteristics were the most salient correlates of a group in which mothers reported more problems than their children.

Determinants of mother-child reporting discrepancies

We examined mother-child reporting discrepancies by modeling patterns in mother and child reports of problem behavior using latent profile analyses and used actual informant scores instead of a unidimensional discrepancy measure. Four profiles, combining mothers' and children's reports, were identified that could be distinguished by the *magnitude* and the *direction* of discrepancies, and by the relative *level* of child problem behaviors.

Multiple variables were independently related to each of these profiles. Mothers with lower educational levels generally reported more problems than their children. Further, higher maternal age (which is a proxy for higher socio-economic status in the Netherlands) and the presence of a sibling in the household were associated with a group of mothers that reported less problems than

their children. The presence of another child in the family may raise the threshold for mothers to consider her child's behavior as problematic.

Also, the boy-girl ratio differed between each of the profiles. If the child was a girl, problem scores were generally lower and discrepancy between information from children and mothers was less common. In contrast, boys were more often found in the two groups characterized by more discrepancy. Interestingly, this association was independent of the direction of discrepancy. Previous studies yielded null or inconsistent findings regarding the effect of gender on informant discrepancies (Achenbach et al., 1987; Grills & Ollendick, 2003; Kolko & Kazdin, 1993), which made De Los Reyes and Kazdin (2005) suggest that child gender may not be related to informant discrepancies. The present findings suggest that gender determines the magnitude of discrepancy and the severity of problems reported, rather than the direction of discrepancy. Many previous studies were unable to distinguish between severity of problems, and the magnitude and direction of discrepancy in their analyses.

Parental psychopathology was -with the exception of a borderline significant effect of maternal anxiety- not related to mother-child reporting discrepancy in this study. This finding is not consistent with the view that parents with psychopathology overestimate their children's problem behavior (Boyle & Pickles, 1997; Collishaw et al., 2009; Kolko & Kazdin, 1993). Only in univariate analyses did we observe that mothers with psychological problems consistently reported a higher level of problem behavior than their children. We discuss two explanations as to why we did not find a bias attributed to caregiver psychological problems. First, unlike most of the other determinants in this study, we measured parental psychological problems when the child was three years of age, instead of concurrent with the assessments of problem behavior. This time-lag may have influenced the effects we found for parental psychological problems as concurrent high levels of depressive symptoms may influence maternal reports of children's problems more than depressive symptoms three years before. However, reporting tendencies of psychopathological symptoms –rather than specific disorders– are relatively stable across time. Also, previous research in the same cohort has shown that depressive symptoms in mothers were stable across three years' time (Cents et al., 2013). Nevertheless, the effects of depressive, anxiety and hostility symptoms on mother-child reporting discrepancies may have been underestimated in the present study as a consequence of the absence of concurrent measures of psychopathology. Second, we controlled the analyses for family functioning as it is strongly associated with parental psychopathology (e.g. Dickstein et al., 1998; Low & Stocker, 2005; Velders et al., 2011). Other studies that took the interdependency of depression and family functioning into account also reported no effect of parental psychological problems on informant discrepancies, or found that effects were greatly reduced (De Los Reyes et al., 2011; Youngstrom et al., 2000). To some extent controlling for interdependent factors may be considered overadjustment. Family functioning and parental psychopathology are closely intertwined and may determine each other.

Family and parenting characteristics were independently associated with informant discrepancies. We found few mother-child reporting discrepancies in those families with less

problems in family functioning, or low levels of harsh discipline. A stable family environment may support communication, mutual understanding (Grills & Ollendick, 2003), and a safe environment for children to disclose their feelings (Goodman et al., 2010). In contrast, more problems in family functioning and lower maternal education were associated with mothers reporting higher problem levels than their children. An unstable family environment may interfere with mothers' ability to cope with a child's behavior and may result in a low threshold for reporting problematic behavior (De Los Reyes & Kazdin, 2005). Alternatively, children who grow up in a more problematic family environment, may not be used to expressing their feelings. Possibly, such children fear the consequences of disclosing emotional topics. Both processes may explain why more problems in family functioning were associated with a group of mothers that reported more problems than their children. Yet, alike parental psychopathology, parenting practices were assessed on average three years before the assessment of mother-child reporting discrepancies. Although parenting styles are moderately stable across time (Huang et al., 2009), effects may have been stronger if had assessment of parenting styles been concurrent

When interpreting associations between parental psychopathology, family and parenting characteristics and informant discrepancies, it must be taken in to account that (except for father reported variables) the mothers provided information on both the determinants and on one of the outcomes included in the discrepancy profiles. If one source provides information on both determinant and outcome, associations may be inflated due to shared-method variance (Collishaw et al., 2009): some portion of the variance between the determinant and the outcome is likely to be shared as result of specific characteristics of the informant (e.g. the informants' perspective on the world, test taking attitudes) rather than as a result of the specific determinant influencing the outcome (LaGrange & Cole, 2008). However, the associations we found are unlikely to completely reflect such spurious inflations. First, maternal psychopathology was not associated with any of the discrepancy profiles. Second, harsh discipline reported by fathers and harsh discipline reported by the mother were similarly related to the discrepancy profiles. Third, associations between observed maternal sensitivity and mother-child reporting discrepancies further supported the validity of the analyses using maternal reported data: more observed maternal sensitivity was associated with a higher likelihood of being included in a profile in which mothers and children reported the lowest problem scores. Finally, more mother-child reporting discrepancies were related to higher teacher reported problem scores, indicating that if either mothers or children perceived problems, these were also apparent for an independent observer. These associations with teacher reported problems support the validity of our profiles.

Children with lower IQ and lower verbal ability were more likely to disagree with their mothers about problem behavior. Previous studies that have taken IQ into account did not find such a relation, however, these studies were all conducted in smaller, clinical, samples (Chi & Hinshaw, 2002; De Los Reyes & Kazdin, 2006; Penney & Skilling, 2012). As a result of the smaller sample sizes, the power to detect an independent effect of IQ may have been limited. In addition, in the study by Penney and Skilling (2012) there was little variance in the IQ range as most adolescents had a relatively low IQ. In the present study, child cognitive abilities were a determinant of membership

in the two profiles with the largest mother-child reporting discrepancies. Interestingly, cognitive difficulties were one of many determinants of the group of mothers reporting more problems than their children. At the same time, cognitive difficulties were a more specific determinant of the group of children reporting more problems than their mothers. Several mechanisms may explain why lower child cognitive abilities determined discrepancies. First, in general, children with a lower IQ and verbal difficulties could, despite the age-appropriate instrument used, have more trouble to report their emotions and behavior. This can result in either under or over reporting of problems. Indeed, a study on credibility of informants showed that clinicians judged children with lower cognitive abilities to be less credible informants (Youngstrom et al., 2011). Second, children with a lower IQ and verbal difficulties may really suffer from more problems, in particular in a cognitive environment like for example school. Some of the mother-child reporting discrepancies elucidated by this study may inform us about problems in a context that mothers cannot easily observe. Indeed, additional analyses with teacher reported data showed that teachers also reported more child problems if children reported more problems than their mothers.

Methodological considerations

Two methodological issues are important to discuss. First, discrepancies can only exist if problems are reported by at least one or the other informant. Thus, determinants of informant discrepancies and determinants of the level of problem behavior may be closely related. Indeed, we found that the profiles with larger discrepancies also had the highest average problem scores. Yet, the average level of problems could not fully explain our results as the two profiles with highest problem levels, but with a different direction of discrepancies in mother and child reports, had very distinct determinants. Second, we studied mother-child reporting discrepancies using non-parallel assessment methods: a puppet interview versus a questionnaire with a different scale structure and different items. This difference in instruments contributed per definition to the discrepancies. In other words, the study design introduced some random error by the choice of measurement tools (Kraemer et al., 2003). Nevertheless, there were several specific determinants clearly associated with each profile. It is unlikely that these specific associations only explained random error. Rather, the findings of the present study are arguably ecologically valid, because different assessments tools reported by different informants are often used in research and clinical settings.

Strengths and limitations

This study has considerable strengths, including the use of young children's self-reports of their problems thereby expanding the age range of previous studies, the large sample size which allowed examining the unique contribution of multiple determinants, and the comprehensive set of characteristics that were jointly studied in association with informant discrepancies. Still, several limitations need to be discussed. First, we aimed at assessing determinants that influenced mother-child reporting discrepancies. However, we did not obtain concurrent measures of parental psychopathology and parental harsh discipline. Although previous studies indicated, as discussed before, that psychopathological symptoms and parenting styles are rather stable over time, the

time-lag may have influenced our results. Yet, this time-lag may rather have led to underestimation of associations than to overestimation. Second, the entropy of the latent profile solution was .68, indicating that there was some insecurity in the estimation of the profiles. Consequently, the small effects of some of the determinants must be interpreted cautiously. Third, our non-response analysis demonstrated that participants included in the study had a higher socioeconomic status and were more often of Dutch national origin. Therefore, one should be careful when generalizing these findings. Yet, selective participation in cohort studies particularly constitutes a problem for estimating prevalence, but less likely biases the associations between variables (Wolke et al., 2009).

Conclusions and implications

The current study demonstrates that determinants of mother-child reporting discrepancies largely depend on the relative problem levels and, importantly, on which informant reports a higher level of problems. For instance, a positive family environment is associated with the absence of discrepancies, while an adverse family environment increases the likelihood that mothers report more problems than their children. Alternatively, child-related characteristics and cognitive difficulties in particular increase the likelihood that children report more problems than their mothers. These findings have an important implication for the way discrepancies in maternal and self-reports of young child problem behavior are perceived. If a mother reports more problems than her child, researchers or clinicians should be aware that family conflict, in addition to other environmental factors and child cognitive functioning, may underlie these discrepancies. In contrast, when a child reports much more problems than his or her mother, one should consider that this could mainly be determined by child related characteristics such as a child's cognitive functioning.

The clinical utility of our results should, however, be evaluated taking into account the following issues. We examined children below the age of eight, different processes may play a more or less important role in older children. In addition, even though we identified specific determinants of mother-child reporting discrepancies, effect estimates were small. Nevertheless, the associations observed in this large non-clinical sample may represent important mechanisms on an individual level, particularly in clinical settings.

Discrepancies between mothers' and young children's reports of problem behavior point to valuable information about underlying reporter characteristics. At the same time, knowledge of families and children can facilitate interpreting discrepancies in mother-child reports that are often difficult to conceptualize and can thereby facilitate diagnostic decisions.

SUMMARY

Informant discrepancies are common when multiple informants rate child or adolescent problems. Discrepancies can influence how researchers and clinicians interpret information, and may thus impact decision making in research and clinical practice. In this study the unique contribution of multiple determinants to discrepancies in mothers' and young children's reports of problem

behavior was examined. Discrepancies were examined by modeling patterns in maternal and children's reports of child internalizing and externalizing problems using latent profile analysis. Mother reported child problems were obtained with the Child Behavior Checklist and children's selfreported problems with the Berkeley Puppet Interview in 5,414 6-year-old children. Four profiles were identified that could be distinguished by problem level, and the direction and magnitude of mother–child reporting discrepancies: a first profile represented agreement (46 %), a second profile represented slight discrepancies (30 %), than there were two profiles with higher problem levels and more discrepancies. The latter two profiles were characterized either by the children (11 %) or the mothers (13 %) reporting more problems. Compared to the first profile, the second profile was predominantly characterized by a positive family environment, the third profile by child cognitive difficulties, and the fourth profile by harsh parenting and poor family functioning. This study demonstrated that determinants of discrepancy in mother and child reports of problem behavior largely depend on the level of problems and on which informant reports a higher level of problems. Knowledge of specific child and family related characteristics that may impact mother–child reporting discrepancies can help the interpretation of each informant's report of child problems and can thereby facilitate diagnostic decisions.

REFERENCES

Ablow, J. C., & Measelle, J. R. (2003). *Manual for the Berkeley Puppet Interview: Symptomatology, social, and academic modules (bpi 1.0)*. MacArthur foundation Research Network Psychopathology and Development (David J. Kupfer, Chair): The MacArthur Working Group on Outcome Assessment

Ablow, J. C., Measelle, J. R., Kraemer, H. C., Harrington, R., Luby, J., Smider, N., Dierker, L., Clark, V., Dubicka, B., Heffelfinger, A. M. Y., Essex, M. J., & Kupfer, D. J. (1999). The MacArthur Three-City Outcome Study: Evaluating multi-informant measures of young children's symptomatology. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*(12), 1580-1590.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213-232.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth & Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Berg-Nielsen, T., Solheim, E., Belsky, J., & Wichstrom, L. (2012). Preschoolers' psychosocial problems: In the eyes of the beholder? Adding teacher characteristics as determinants of discrepant parent–teacher reports. *Child Psychiatry and Human Development, 43*(3), 393-413.

Boyle, M. H., & Pickles, A. (1997). Maternal depressive symptoms and ratings of emotional disorder symptoms in children and adolescents. *Journal of Child Psychology and Psychiatry, 38*(8), 981-992.

Briggs-Gowan, M. J., Carter, A. S., & Schwab-Stone, M. (1996). Discrepancies among mother, child, and teacher reports: Examining the contributions of maternal depression and anxiety. *Journal of Abnormal Child Psychology, 24*(6), 749-765.

Cents, R. A. M., Diamantopoulou, S., Hudziak, J. J., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., Lambregtse-van den Berg, M. P., & Tiemeier, H. (2013). Trajectories of maternal depressive symptoms predict child problem behaviour: The Generation R Study. *Psychological Medicine, 43*(01), 13-25.

Chi, T. C., & Hinshaw, S. P. (2002). Mother-child relationships of children with adhd: The role of maternal depressive symptoms and depression-related distortions. *Journal of Abnormal Child Psychology, 30*(4), 387-400.

Collishaw, S., Goodman, R., Ford, T., Rabe-Hesketh, S., & Pickles, A. (2009). How far are associations between child, family and community factors and child psychopathology informant-specific and informant-general? *Journal of Child Psychology and Psychiatry, 50*(5), 571-580.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16*(3), 330-334.

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483-509.

De Los Reyes, A., & Kazdin, A. E. (2006). Informant discrepancies in assessing child dysfunction relate to dysfunction within mother-child interactions. *Journal of Child and Family Studies, 15*(5), 643-661.

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles underlying the use of multiple informants' reports. *Annual review of clinical psychology, 9*, 123-149.

De Los Reyes, A., Youngstrom, E. A., Pabon, S. C., Youngstrom, J. K., Feeny, N. C., & Findling, R. L. (2011). Internal consistency and associated characteristics of informant discrepancies in clinic referred youths age 11 to 17 years. *Journal of Clinical Child and Adolescent Psychology, 40*(1), 36-53.

Derogatis, L. R., editor. (1993). The Brief Symptom Inventory (BSI): Administration, scoring, and procedures. Manual. 3rd ed. Minneapolis, MN: National Computer System Inc.

Dickstein, S., Seifer, R., Hayden, L. C., Schiller, M., Sameroff, A. J., Keitner, G., Miller, I., Rasmussen, S., Matzko, M., & Magee, K. D. (1998). Levels of family assessment: Ii. Impact of maternal psychopathology on family functioning. *Psychological Assessment, 12*(1), 23-40.

Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice, 7*(4), 435-453.

Frick, P. J., Silverthorn, P., & Evans, C. (1994). Assessment of childhood anxiety using structured interviews: Patterns of agreement among informants and association with maternal anxiety. *Psychological Assessment, 6*(4), 372-379.

Goodman, K., De Los Reyes, A., & Bradshaw, C. (2010). Understanding and using informants' reporting discrepancies of youth victimization: A conceptual model and recommendations for research. *Clinical Child and Family Psychology Review, 13*(4), 366-383.

Grietens, H., Onghena, P., Prinzie, P., Gadeyne, E., Van Assche, V., Ghesquière, P., & Hellinckx, W. (2004). Comparison of mothers', fathers', and teachers' reports on problem behavior in 5- to 6-year-old children. *Journal of Psychopathology and Behavioral Assessment, 26*(2), 137-146.

Grills, A. E., & Ollendick, T. H. (2002). Issues in parent-child agreement: The case of structured diagnostic interviews. *Clinical Child and Family Psychology Review, 5*(1), 57-83.

Grills, A. E., & Ollendick, T. H. (2003). Multiple informant agreement and the anxiety disorders interview schedule for parents and children. *Journal of the American Academie of Child and Adolescent Psychiatry, 42*(1), 30-40.

Huang, K.-Y., O'Brien Caughy, M., Lee, L.-C., Miller, T., & Genevro, J. (2009). Stability of maternal discipline practices and the quality of mother–child interaction during toddlerhood. *Journal of Applied Developmental Psychology, 30*(4), 431-441.

Jaddoe, V. W., van Duijn, C. M., Franco, O. H., van der Heijden, A. J., van IJzendoorn, M. H., de Jongste, J. C., van der Lugt, A., Mackenbach, J. P., Moll, H. A., Raat, H., Rivadeneira, F., Steegers, E. A., Tiemeier, H., Uitterlinden, A. G., Verhulst, F. C., & Hofman, A. (2012). The Generation R Study: Design and cohort update 2012. *European Journal of Epidemiology, 27*(9), 739-756.

Jansen, P. W., Raat, H., Mackenbach, J. P., Hofman, A., Jaddoe, V. W. V., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Verhulst, F. C., & Tiemeier, H. (2012). Early determinants of maternal and paternal harsh discipline: The Generation R Study. *Family Relations, 61*(2), 253-270.

Kok, R., Linting, M., Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2013). Maternal sensitivity and internalizing problems: Evidence from two longitudinal studies in early childhood. *Child Psychiatry and Human Development, 44*(6), 751-765.

Kolko, D. J., & Kazdin, A. E. (1993). Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent and teacher reports. *Journal of Child Psychology and Psychiatry, 34*(6), 991-1006.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*(9), 1566-1577.

Krain, A. L., & Kendall, P. C. (2000). The role of parental emotional distress in parent report of child anxiety. *Journal of Clinical Child Psychology, 29*(3), 328-335.

LaGrange, B., & Cole, D. A. (2008). An expansion of the trait-state-occasion model: Accounting for shared method variance. *Structural Equation Modeling, 15*(2), 241-271.

Lavigne, J., Dahl, K., Gouze, K., LeBailly, S., & Hopkins, J. (2014). Multi-domain predictors of oppositional defiant disorder symptoms in preschool children: Cross-informant differences. *Child Psychiatry and Human Development*, 1-12.

Low, S. M., & Stocker, C. (2005). Family functioning and children's adjustment: Associations among parents' depressed mood, marital hostility, parent-child hostility, and children's adjustment. *Journal of Family Psychology, 19*(3), 394-403.

Miller, I. W., Epstein, N. B., Bishop, D. S., & Keitner, G. I. (1985). The McMaster Family Assessment Device: Reliability and validity. *Journal of Marital and Family Therapy, 11*(4), 345-356.

Muthén, L. K., & Muthén, B. (1998–2010). *Mplus user's guide, 6th*. . Los Angeles, CA: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569.

Penney, S. R., & Skilling, T. A. (2012). Moderators of informant agreement in the assessment of adolescent psychopathology: Extension to a forensic sample. *Psychological Assessment, 24*(4), 386-401.

Reid, J. B., Kavanagh, K., & Baldwin, D. V. (1987). Abusive parents' perceptions of child problem behaviors: An example of parental bias. *Journal of Abnormal Child Psychology, 15*(3), 457-466.

Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin, 112*(3), 485-499.

Ringoot, A. P., Jansen, P. W., Steenweg-de Graaff, J., Measelle, J. R., van der Ende, J., Raat, H., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2013). Young children's self-reported emotional, behavioral, and peer problems: The Berkeley Puppet Interview. *Psychological Assessment, 25*(4), 1273-1285.

Rubin, D. H., Althoff, R. R., Walkup, J. T., & Hudziak, J. J. (2013). Cross-informant agreement on child and adolescent withdrawn behavior: A latent class approach. *Child Psychiatry and Human Development, 44*(3), 361-369.

Statistics Netherlands. (2004). Allochtonen in Nederland 2004 [Foreigners in the Netherlands 2004]. Voorburg/ Heerlen.

Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of child maltreatment with the parent-child conflict tactics scales: Development and psychometric data for a national sample of American parents. *Child Abuse and Neglect, 22*(4), 249-270.

Strickland, J., Hopkins, J., & Keenan, K. (2012). Mother-teacher agreement on preschoolers' symptoms of ODD and CD: Does context matter? *Journal of Abnormal Child Psychology, 40*(6), 933-943.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (2005). *Snijders-Oomen niet-verbale intelligentietest SON-r 2, 5-7*. Handleiding [Snijders-Oomen nonverbal intelligence test SON-r 2, 5-7. Manual]. Amsterdam: Boom Testuitgevers.

Treutler, C. M., & Epkins, C. C. (2003). Are discrepancies among child, mother, and father reports on children's behavior related to parents' psychological symptoms and aspects of parent-child relationships? *Journal of Abnormal Child Psychology, 31*(1), 13-27.

Valla, J. P., Bergeron, L., & Smolla, N. (2000). The Dominic-R: A pictorial interview for 6- to 11-year-old children. *Journal of the American Academy of Child & Adolescent Psychiatry, 39*(1), 85-93.

Van Bon, W. H. J., & Hoekstra, J. G. (1982). *Taaltests voor kinderen*. Lisse: Swets & Zeitlinger.

van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment, 24*(2), 293-300.

Velders, F. P., Dieleman, G., Henrichs, J., Jaddoe, V. W., Hofman, A., Verhulst, F. C., Hudziak, J. J., & Tiemeier, H. (2011). Prenatal and postnatal psychological symptoms of parents and family functioning: The impact on child emotional and behavioural problems. *European Child and Adolescent Psychiatry, 20*(7), 341-350.

Verhulst, F. C., & van der Ende, J. (2008). Using rating scales in a clinical context. In M. Rutter, D. V. M. Bishop, D. S. Pine, S. Scott, J. Stevenson, E. Taylor & A. Thapar (Eds.), *Rutter's child and adolescent psychiatry, 5th edition*: Blackwell Publishing Limited.

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., & Lamberts, K. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *British Journal of Psychiatry, 195*(3), 249-256.

Youngstrom, E., Izard, C., & Ackerman, B. (1999). Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology, 67*(6), 905-916.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 168*, 1038-1050.

Youngstrom, E., Youngstrom, J. K., Freeman, A. J., De Los Reyes, A., Feeny, N. C., & Findling, R. L. (2011). Informants are not all equal: Predictors and correlates of clinician judgments about caregiver and youth credibility. *Journal of Child and Adolescent Psychopharmacology, 21*(5), 407-415.

SUPPLEMENTARY MATERIAL

Figure. S1 Graphical representation of the subject flow

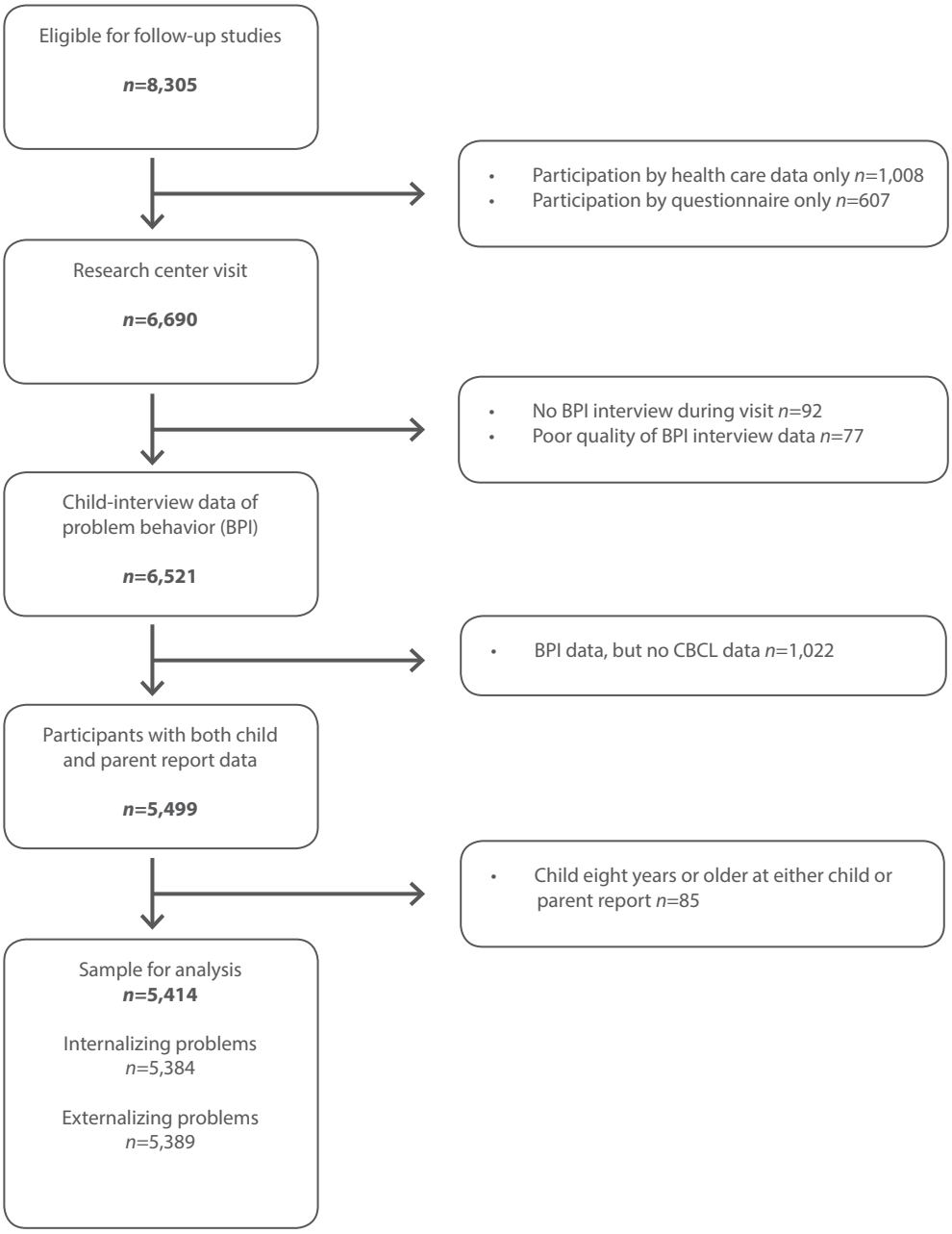


Table S1. Unadjusted associations of potential determinants of disagreement with profile membership

	Child= Mother *(1st profile)* *n* = 2,474), *Ref.*	Child > Mother *(2nd profile)* *n* = 1,634			Child >> Mother *(3rd profile)* *n* =6 14			Child << Mother *(4th profile)* *n* = 692		
	Mean / %[a]	*M* / %[a]	*OR*	95%*CI*	*M* / %[a]	*OR*	95%*CI*	*M* / %[a]	*OR*	95%*CI*
Demographic and Socioeconomic Characteristics										
Maternal age at pregnancy	31.1	31.6	1.02**	1.01;1.04	30.6	0.98*	0.96;0.99	30.0	0.96***	0.94;0.97
Child age	6.09	6.07	0.99	0.98;1.01	6.1	1.01	0.99;1.03	6.1	1.02	0.99;1.04
Child Gender										
Boy	47.9	43.3	0.83**	0.73;0.94	65.1	2.04***	1.69;2.45	58.4	1.53***	1.29;1.81
Girl	52.1	56.7	1		34.9	1		41.6	1	
Siblings										
No	19.4	15.1	0.74***	0.62;0.88	21.8	1.15	.92;1.44	23.5	1.27*	1.03;1.56
Yes	80.6	84.9	1		78.2	1		76.5	1	
Ethnicity										
Non-Western	30.5	25.3	0.77***	0.67;0.89	40.1	1.57***	1.33;1.85	47.5	2.11***	1.77;2.52
Other-Western	8.4	9.1	1.01	0.80;1.26	8.8	1.26	0.91;1.73	7.4	0.29	0.86;1.66
Dutch	61.1	65.6	1		51.1	1		45.1	1	
Educational level										
Primary	3.3	3.1	0.85	0.59;1.23	4.6	1.60*	1.02;1.51	9.5	3.94***	2.78;5.60
Secondary	39.1	34.0	0.80***	0.70;0.91	45.2	1.33**	1.11;1.60	39.3	1.66***	1.39;1.98
High	57.6	63.0	1		50.2	1		56.5	1	
Marital Status										
Single parenthood	13.2	10.1	0.74**	0.61;0.91	17.9	1.44**	1.14;1.83	20.9	1.74***	1.40;2.16
Married/cohabiting	86.8	89.9	1		82.1	1		79.1	1	

Table S1. continued

Income										
<1200	6.6	4.3	0.54***	0.40;0.73	9.7	2.02***	1.42;2.88	13.2	3.15***	2.24;4.43
1200-2000	14.8	12.9	0.72**	0.59;0.88	21.5	1.98***	1.51;2.58	24.5	2.57***	1.96;3.36
2000-4000	44.2	41.4	0.78***	0.68;0.90	43.7	1.36**	1.09;1.69	40.3	1.44**	1.14;1.81
>4000	34.4	41.4	1		25.1	1		22.0	1	
Parental Psychopathology [b,c]										
Depression mother	0.15	0.10	0.82***	0.74;0.91	0.19	1.14**	1.04;1.25	0.27	1.29***	1.18;1.40
Anxiety mother	0.20	0.15	0.80***	0.73;0.88	0.23	1.10*	1.01;1.19	0.33	1.31***	1.21;1.41
Hostility mother	0.21	0.15	0.76***	0.68;0.84	0.23	1.08	0.98;1.19	0.31	1.28***	1.19;1.39
Depression father	0.14	0.10	0.88**	0.81;0.96	0.16	1.07	0.99;1.17	0.20	1.13**	1.05;1.22
Anxiety father	0.20	0.16	0.84***	0.78;0.91	0.22	1.07	0.97;1.18	0.27	1.18***	1.09;1.28
Hostility father	0.21	0.17	0.85***	0.77;0.92	0.23	1.09	0.97;1.22	0.27	1.17***	1.08;1.26
Family and Parenting [b,c]										
Adverse family functioning	1.5	1.4	0.39***	0.33;0.46	1.6	1.38**	1.12;1.70	1.7	2.94***	2.42;3.58
Harsh discipline mother	2.3	1.8	0.82***	0.79;0.86	2.7	1.10***	1.05;1.16	3.2	1.20***	1.14;1.25
Harsh discipline father	1.9	1.5	0.87***	0.83;0.91	2.4	1.12***	1.06;1.19	2.6	1.16***	1.10;1.22
Cognitive abilities of the child [d]										
IQ	102.9	104.1	1.01**	1.00;1.01	97.4	0.97***	0.97;0.98	98.1	0.98***	0.97;0.98
Language [b]	22.25	22.47	1.03*	1.01;1.06	21.15	0.89***	0.87;0.92	21.40	0.91***	0.89;0.94

[a] Presented are raw and un-adjusted scores [b] Odds ratios are presented 'per *SD*' [c] Higher values represent more problems [d] Higher values represent fewer problems, *$p < .05$, **$p < .01$, ***$p < .001$

chapter 5

Parental depression and child well-being: young children's self-reports helped addressing biases in parent reports

Ringoot, A.P., Tiemeier, H., Jaddoe, V.W.V., So, P., Hofman, A., Verhulst, F.C., & Jansen, P.W.

Journal of Clinical Epidemiology. Accepted for publication.

ABSTRACT

Objectives: Effects of maternal and paternal depression on child development are typically evaluated using parental reports of child problems. Yet, parental reports may be biased.

Methods: In a population-based cohort, parents reported lifetime depression (n = 3,178) and depressive symptoms (n = 3,131). Child emotional and behavioral problems were assessed at age 6 years by child self-report using the Berkeley Puppet Interview, and by mother report using the Child Behavior Checklist (CBCL), and at age 3 years by father and mother reported CBCLs.

Results: Both maternal and paternal depression was associated with more child problems. Associations were of similar strength if child problems were obtained by self-reports. However, if parents reported about their own depression or depressive symptoms and about their child's problems, estimates were generally stronger for associations with the reporting parent's depression as the determinant. For instance, if mothers reported child emotional problems, associations were stronger for maternal (B = 0.27, 95% CI = 0.19; 0.35) than for paternal lifetime depression (B = 0.12, 95% CI = 0.02; 0.21, p-value for difference = .02).

Conclusion: Depression of mothers and fathers affects young children's well-being. However, if parents reported about their own depression and about child problems, associations were inflated. To accurately estimate effects of parental depression, multiple-source data including young children's perspectives must be considered.

INTRODUCTION

Parental depression places children at risk for a range of adverse outcomes including substance dependence, social problems and poor physical health later in life (Connell & Goodman, 2002; Weissman et al., 2006). While the risk that maternal depression poses for offspring's unhealthy development has often been studied (Goodman & Gotlib, 1999), the effect of paternal depression has only recently received more attention (Ramchandani & Psychogiou, 2009). As fathers not only pass on their genes, but are increasingly involved in care-giving activities (Cabrera, Tamis-LeMonda, Bradley, Hofferth, & Lamb, 2000), it is important to examine the impact of both maternal and paternal depression on children's development. This is particularly important in early childhood, as the first years of life are critical for the cognitive, physical and emotional development of children. Children rely heavily on their parents in the preschool period, as they are children's primary environment.

Evidence of the specific impact of fathers' depression compared to mothers' depression is relatively scarce. Moreover, our knowledge about the effect of maternal and paternal depression on offspring development in early life is mainly based on parent –mostly mother– reports of child problems (Kvalevaag et al., 2013; Malmberg & Flouri, 2011; Ramchandani, Stein, Evans, & O'Connor, 2005; Velders et al., 2011; Weitzman, Rosenthal, & Liu, 2011). This potentially limits conclusions of such studies: if the same source is used to provide information on both determinant and outcome, associations may be inflated due to shared-method variance (Collishaw, Goodman, Ford, Rabe-Hesketh, & Pickles, 2009). In the case that the same informant reports both on the determinant and the outcome, some portion of the variance between the determinant and the outcome is likely to be shared as result of specific characteristics of the informant (e.g. the informants' perspective on the world, test taking attitudes) rather than as a result of the construct of interest (LaGrange & Cole, 2008). Moreover, depressed parents may overestimate their children's problems (Richters, 1992).

Only a few studies on the impact of maternal and paternal depression complemented parent reports of young children's problems with information from other sources, like teachers (Gere et al., 2013; Ramchandani et al., 2008a; Ramchandani et al., 2008b). However, these studies generally used mean scores or generated combined diagnoses, thus diagnosing a problem if it was present according to either one of the informants. As a result, bias by parental report could not be ruled out.

In current literature, one potential source of information is largely ignored: the child's own perspective on his or her problems. To the best of our knowledge, child report has not been used as a source when examining the association between parental depression and young children's problem behavior. The offspring's perspective on well-being has only been considered in studies assessing the effect of parental psychopathology among adolescents. Interestingly, some of these studies showed that fathers' depressive symptoms *during* the adolescent period were more strongly associated with emotional and behavioral problems than mothers' depressive symptoms (Compas, Phares, Banez, & Howell, 1991; Papp, 2012). In contrast, weaker associations of paternal rather than maternal *perinatal* depression with adolescent self-reported outcomes have been observed (Pearson et al., 2013), suggesting that early in life the impact of maternal depression may be larger (Connell & Goodman, 2002).

Unfortunately, similar studies using young children's own reports of their problems to compare the impact of maternal and paternal depression are not available. Obtaining self-reports of problems in children under the age of eight has been considered challenging given the relative lack of appropriate instruments. However, studies have shown that by using an age-appropriate method, five-to-seven year old children from diverse socio-demographical backgrounds can provide valuable information on their problems (Ablow et al., 1999; Arseneault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005; Luby, Belden, Sullivan, & Spitznagel, 2007; Ringoot et al., 2013).

The present study compared the association of paternal and maternal depression with emotional and behavioral problems of three-to-six-year-old children using father, mother and self-reports of child problems. We hypothesize that by using information about child problem behavior from more than one informant, a more accurate picture of the effect of parental depression is obtained than by relying on mother or father reports of child problems only. Given results from previous studies (Connell & Goodman, 2002), we also hypothesize that fathers' depression has an adverse but smaller effect on child problems than mothers' depression. To examine consistency of associations throughout different phases in young children's development, we studied lifetime depression before childbirth and parental depressive symptoms when the child was three years of age.

METHODS

Design and Study Population

This study was performed within Generation R, a population-based prospective cohort from fetal life onwards in Rotterdam, the Netherlands (Jaddoe et al., 2012). The study was approved by the Medical Ethics Committee at Erasmus University Medical Center Rotterdam. Written informed consent was obtained from all caregivers.

The analyses for this study were performed in two subsamples from the Generation R Study. The first subsample was defined by families with complete information on maternal and paternal lifetime depression, assessed prenatally (n = 4,545). Analyses were performed using mother and father reported child outcomes at age 3 years, and using mother and child self-reported outcomes at age 6 years. Of the 4,545 participants with maternal and paternal prenatally assessed lifetime depression data, 3,582 participants had outcome data at either age 3 or 6 years. Of these, 2,573 participants had data on mother and father reported child outcomes at age 3 years, and 3,178 participants had data on mother and child reported child outcomes at age 6 (flowchart in supplementary material, Figure S1).The second subsample was defined by the availability of maternal and paternal reports of depressive symptoms, assessed at child age 3 years (n = 3,718). Again, analyses were performed with child outcomes at child age 3 and child age 6 years. Of the 3,718 participants with maternal and paternal depressive symptom data assessed at child age 3 years, 3,660 participants had data on mother and father reported child outcomes at age 3 years, and 3,131 participants had data on mother and child reported child outcomes at age 6 years (Figure S2).

The overlap between the two subsamples was 2,091 (i.e. all determinants and outcome data at age 6 years), thus 1,087 participants were included in the first, but not in the second subsample. These participants were of more disadvantaged socioeconomic background (see non-response analyses).

Non-response analyses

Two non-response analyses were performed. The first compares participants included in both subsamples (n = 2,091) with participants in either one of the two subsamples (n = 1,087). The second is a loss to follow-up analysis, comparing participants included in the analyses of prenatal depression and child outcomes at age 6 years (n = 3,178) with participants that had baseline data, but were lost to follow-up (n = 1,367 (4,545 - 3,178)).

The first non-response analysis indicated that participants included in just one of the two subsamples (n = 1,087) had more often younger mothers (M_1 = 29.7, SD_1 = 4.9; M_2 = 31.8, SD_2 = 3.9, t(1,824) = -11.95, p < .001), were more often of non-Western origin (35% non-Western vs. 13% non-Western, $\chi^2(1) = 209$, p < .001), came from families with a lower (< 2000 € net monthly) household income (38% vs. 15%, $\chi^2(1) = 185$, p < .001), and a lower level (primary or secondary) of parental education (maternal: 59% vs. 33%, $\chi^2(1) = 192$, p < .001; paternal: 60% vs. 37%, $\chi^2(1) = 150$, p < .001), than participants included in both subsamples (n = 2,091).

The second analysis of non-response indicated that participants that were lost to follow-up (n = 1,367) had more often younger mothers (M_1 = 29.1, SD_1 = 5.2; M_2 = 31.1, SD_2 = 4.4, t(2,239) = -12.46, p < .001), were more often of non-Western origin (32% non-Western vs. 20% non-Western, $\chi^2(1) = 72$, p < .001), came from families with a lower (< 2000 € net monthly) household income (38% vs. 22%, $\chi^2(1) = 106$, p < .001), and a lower level (primary or secondary) of parental education (maternal: 57% vs. 42%, $\chi^2(1) = 80$, p < .001; paternal: 57% vs. 45%, $\chi^2(1) = 54$, p < .001), than participants included in the analyses (n = 3,178).

Parental Depression

Lifetime depression was assessed prior to childbirth at 20 weeks pregnancy using self-report questionnaires filled out by mothers and fathers separately. At the time of responding 94.4% of the parents lived in the same household. Parents indicated whether they had ever suffered from psychiatric disorders. For each disorder an example was provided, followed by several items. We used the item on lifetime depression. To validate our measure of self-reported lifetime depression, we compared these reports to a diagnosis of lifetime depression obtained with the Composite International Diagnostic Interview (CIDI; WHO, 1990), in a subsample from the overall cohort. The use of the CIDI in the Generation R Study has previously been described in more detail (Tharner et al., 2012). Self-reported lifetime depression had a sensitivity of 80% and a specificity of 80% for mothers (n = 615), and a sensitivity of 70% and specificity of 87% for fathers (n = 584).

Prevalent depressive symptoms of mothers and fathers at child age 3 years were assessed with the Dutch version of the Brief Symptom Inventory (BSI), a validated self-report questionnaire which is

widely used in clinical and research settings (De Beurs, 2004). Based on self-reported information, 99% of the fathers were the biological father of the child. The original BSI consists of 53 items, scored on a 5-point scale, ranging from 0 to 4. The items assess psychiatric symptoms occurring in the preceding seven days. We calculated a mean score on the 6-item depression scale (α = .80 for both mothers and fathers).

Children's Emotional and Behavioral Problems

Child self-reported emotional and behavioral problems were obtained with the Berkeley Puppet Interview (BPI; Ablow & Measelle, 2003). This interview was assessed at our research center when children were on average 6 years old. During the interview two identical dog hand puppets made opposing statements about themselves and asked children to indicate which statement described him/her best. Each statement was scored on a 7-point scale, with higher scores representing more problems. Items were summed to compute scale scores. We used the 20-item internalizing (i.e. emotional problems, α = .70) and 21-item externalizing (i.e. behavioral problems, α = .76) scales. Within this cohort, we recently demonstrated that 5-to-7 year old children are capable of providing valid and specific information using the BPI (Ringoot et al., 2013).

Parent reported child emotional and behavioral problems were assessed with the Child Behavior Checklist 1½-5 (CBCL; Achenbach & Rescorla, 2000), when children were 3 (*SD* = 0.10) and 6 years old (*SD* = 0.35). At child age 3 years mothers and fathers each completed the CBCL. At child age 6 years the primary caregiver (92% mothers) completed the questionnaire at home, prior to the research center visit. We will refer to this report as 'mother report at 6 years'. Parents rated the occurrence of their children's behavior within the past two months on a 3-point scale. Items were summed to compute scale scores, higher scores represented more problems. We used the 36-item internalizing (emotional problems) and 24-item externalizing (behavioral problems) scales. Good reliability and validity have been reported for the CBCL (Achenbach & Rescorla, 2000). Internal consistencies ranged between .79-.89.

Covariates

Several potential covariates were considered. Information on *parents age at intake,* and *child gender* was obtained from medical records. *Smoking and alcohol use during pregnancy* were assessed by questionnaires during pregnancy. Information on *number of children in the household, child national origin, parental educational level, marital status* and *family income* were obtained by a questionnaire prior to the research center visit. The child's national origin was based on the countries of birth of both parents (Statistics Netherlands, 2004). Marital status was divided in two categories, indicating whether mothers lived with or without a partner. During the research center visit *child age* was recorded and *children's verbal abilities* were assessed using a receptive language subtest (syntax choice task) of a Dutch test battery (Van Bon & Hoekstra, 1982).

Data analysis

The relation between parental depression and child problems was first analyzed using prenatally assessed parental lifetime depression as the determinant and then with continuous parental depressive symptoms at child age 3 years as the determinant. Analyses were conducted separately for paternal and maternal depression. Children's emotional or behavioral problems were studied as continuous outcomes. Analyses were first performed with child outcome data at child age 6 years (child and mother report of child problems using the BPI and CBCL, respectively) and repeated with outcome data at child age 3 years (mother and father report of child problems using the CBCL). To approach a normal distribution of the outcome data, BPI and CBCL scale scores were transformed using inverse and square root transformations. To facilitate effect size comparison, z-scores were calculated for outcomes and continuous determinants.

First, we examined the relationship between parental depression and child problems using mixed models for multiple source data (Goldwasser & Fitzmaurice, 2001). The outcome was child problems reported by two persons, either child (BPI) and mother (CBCL) at child age 6 years, or mother and father (both CBCL) at child age 3 years. In these mixed models (PROC MIXED in SAS 9.3) the outcome was treated as repeated construct. We formally examined whether the strength of associations between parental depression and child problems depended on the informant of the outcome by including an interaction term between informant (child vs. mother or mother vs. father) and parental depression in the models.

Second, if associations depended on the reporter of the outcome (i.e. a significant interaction in the mixed models), we performed linear regression analyses with parental depression as determinant and used mother and self-report of child problems at child age 6 years, or mother and father reports at child age 3 years as separate outcomes. We used a z-test to examine the difference in estimates of father versus mother depression. Regressions were controlled for covariates if they changed the estimates of the unadjusted association by more than 5%. In a next step, to control for any confounding due to depression of the other parent, analyses were controlled for depression of the other parent (in 6.4% of the couples lifetime depression was present in mothers and fathers; the correlation between maternal and paternal depressive symptoms was $r = .25$). Because analyses were performed in separate subsamples, in additional analyses we repeated the linear regression analyses in participants with complete data.

Third, to illustrate the contribution of parents' depression to child problems, we plotted the estimates of having either a father, a mother or two parents with lifetime depression. Analyses of covariance (ANCOVAs) were performed with a categorical indicator of the number of affected parents as a determinant of BPI and CBCL scores.

Multiple imputation was used to account for missing values in covariates. Missing values ranged from < 0.01% for gender to 14.0% for child verbal abilities (mean percentage missing data on covariates: 4.8%). Missing data on the CBCL at age 6 was imputed if CBCL data at either 1.5 or 3 years was available. Coefficients were averaged over 20 imputed datasets. Regression analyses, ANCOVAs and multiple imputation were performed using SPSS version 20.0 (IBM, Armonk, NY).

Table 1. Participant characteristics in original data

	Subsample with data on prenatal lifetime depression and child problems (*n* = 3,178)		Subsample with data on depressive symptoms at child age 3 years and child problems (*n* = 3,131)	
	n	Mean *(SD)*	*n*	Mean *(SD)*
Parental characteristics				
Paternal age at intake, yr	3,177	33.5 (5.32)	2,527	34.1 (5.00)
Maternal age at intake, yr	3,178	31.1 (4.41)	3,131	32.0 (4.20)
Education father (% Lower)	2,892	35.2	2,936	30.5
Education mother (% Lower)	3,170	36.6	3,124	30.3
Marital status (% Living without partner)	3,008	8.8	2,995	5.7
Household income (%)	2,864		2,884	
<2,000		14.9		9.8
2,000-4,000		44.8		44.0
>4,000		40.3		46.2
Alcohol use throughout pregnancy (% any)	2,905	47.6	2,531	52.1
Smoking during pregnancy (% continued)	2,924	13.2	2,849	10.6
Lifetime depression prenatally (% yes)				
Father	3,178	16.0 (*n* = 510)	-	-
Mother	3,178	27.8 (*n* = 884)	-	-
Depressive symptoms at age 3 (BSI scores)				
Father	-	-	3,131	0.11 (0.27)
[median (90 percentile range)]				[0.00 (0.00, 0.50)]
Mother	-	-	3,131	0.09 (0.26)
[median (90 percentile range)]				[0.00(0.00, 0.50)]
Child characteristic				
Gender (% boys)	3,178	48.6	3,131	49.4
Siblings (% yes)	2,957	83.1	2,969	84.8
Age at BPI assessment, yr	3,178	6.1 (0.35)	3,131	6.0 (0.29)
National origin (%)	3,173		3,128	
Non-Western		20.2		15.8
Other-Western		9.0		9.2
Dutch		70.8		75.0
Verbal ability score	2,736	22.4 (2.81)	2,682	22.4 (2.76)
Child reported problems (BPI 6 yr)				
Emotional problems score	3,178	58.2 (11.97)	3,131	57.1 (11.43)
Behavioral problems score	3,168	51.8 (10.27)	3,121	51.4 (10.05)
Mother reported child problems (CBCL 6 yr)				
Emotional problems score	3,009	5.5 (5.39)	3,005	5.1 (4.97)
Behavioral problems score	3,018	7.0 (6.19)	3,008	6.7 (5.96)

*Note***.** Values represent original, non-imputed data. Presented are mean and standard deviation, unless otherwise indicated.

RESULTS

Characteristics of the study population are presented in Table 1. The mean parental depression scores at child age three years in this sample were lower (0.11 for mothers [*SD* = 0.27]; 0.09 for fathers [*SD* = 0.26]) than in a Dutch norm population (De Beurs, 2004). In this sample, 8% of mothers and 11% of fathers scored above average. About half of the children were boys, about 70% were of Dutch national origin. Correlations between child and parent reported emotional and behavioral problems at 6 years were r = .13 and r = .18, respectively. Correlations between maternal and paternal report of child problems at 3 years were r = .48 for emotional and r = .53 for behavioral problems.

Mixed model analyses showed several interactions between parents' depression and the informant of child problems, indicating that the size and direction of associations between parental depression and child problems depended on whether children, mothers or fathers reported child problems (see supplementary material, Table S1). Since associations depended on the informant, in the next analyses we examined associations between parental depression and child problems for each informant (child, mother and father) separately.

Parental lifetime depression assessed prenatally and child problems at two time points

Table 2 shows that if emotional and behavioral problems were *reported by 6-year-old children,* lifetime depression of fathers and mothers was associated with child *emotional* problems (father B = 0.13, p < .01; mother B = 0.12, p < .01). Only fathers', but not mothers', lifetime depression was associated with child reported *behavioral* problems (father B = 0.16, p < .01; mother B = 0.06, p = .11). If *mothers reported the problems of their 6-year-old children,* results differed from those obtained using child self-reports. The associations of fathers' depression with children's problems were *smaller* than associations of mothers' depression with child problems (e.g. emotional problems: father B = 0.12, p < .05; mother B = 0.27, p < .001). Additional z-tests demonstrated that when child problems were reported by the mother, the difference between maternal and paternal depression reached significance for emotional, but not for behavioral, problems (p-values for difference .02 and .16, respectively). Thus, mothers' depression was more strongly associated with child problems than fathers' depression if mothers reported on child problems, but this difference was not present if children self-reported their problems. A similar, but non-statistically significant, difference was observed when *father and mother reported child problems at child age 3 years* were assessed as the outcome: fathers' lifetime depression was more strongly associated with child problems than mothers' lifetime depression if fathers reported on both their own depression and on child problems, and vice versa. Estimates that are potentially affected by bias are highlighted in Table 2 .

Table 2. Parental prenatal lifetime depression and child problems

	Child age 6 years				Child age 3 years			
	Child report (BPI)[□]		Mother report (CBCL)[□]		Father report (CBCL)[□]		Mother report (CBCL)[□]	
Lifetime depression, prenatally (yes/no)	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*
	Emotional problems score n = 3,178				*Emotional problems score n = 2,573*			
Father	0.13	0.03;0.22**	0.12#	0.02;0.21*	0.25	0.14;0.35***	0.14	0.04;0.25**
Mother	0.12	0.04;0.19**	0.27#	0.19;0.35***	0.17	0.09;0.26***	0.25	0.16;0.33***
	Behavioral problems score n = 3,168				*Behavioral problems score n = 2,573*			
Father	0.16	0.06;0.25**	0.14	0.05;0.24**	0.24	0.13;0.34***	0.18	0.07;0.28**
Mother	0.06	-0.01;0.14	0.23	0.15;0.31***	0.17	0.08;0.26***	0.20	0.11;0.28***

Note. Models adjusted for: parental age, income, marital status, maternal educational level, smoking during pregnancy, child national origin, child gender, child age, number of siblings and verbal abilities of the child. Shades indicate estimates where determinant and outcome were reported by the same informant.

[□] BPI scores inverse transformed, CBCL scores square root transformed. All z-scores.

Significant difference in estimates for paternal and maternal depression.

*$p < .05$, **$p < .01$, ***$p < .001$

Parental depressive symptoms at child age 3 years and child problems at two time points

Table 3 shows that maternal and paternal depressive symptoms at child age 3 years were associated with child emotional (according to parental reports) and behavioral problems (according to all informants).

Similar to the effects of parental lifetime depression, if emotional and behavioral problems were *reported by 6-year-old children,* effects of fathers' versus mothers' depressive symptoms on child problems were of a similar strength (e.g. behavioral problems: father $B = 0.04$, $p < .05$; mother $B = 0.05$, $p < .01$).

In contrast, if child problems were *reported by the mother,* observed effects of paternal depressive symptoms were weaker than those of maternal depressive symptoms (e.g. behavioral problems at age 6: father $B = 0.11$, $p < .001$; mother $B = 0.16$, $p < .001$). In contrast, analyses with *parental reports at child age 3 years* showed that fathers' depressive symptoms were more strongly related to child problems if these were also *reported by fathers*. Differences in estimates between associations of maternal and paternal depression with child outcomes were all significant (*p*-values for difference < .05).

We reran all analyses additionally adjusting for depression of the other parent. This attenuated results: the association of fathers' depression and mother reported child emotional problems and the association of fathers' depressive symptoms and child reported behavioral problems were no longer statistically significant.

In addition, analyses were repeated in participants with complete data (i.e. n = 2,091; supplementary material, Table S2 and Table S3). The results were similar to those presented in Table 2 and 3.

We also conducted these analyses using the non-imputed data only. Results were essentially unchanged: i.e. none of the differences in effect estimates between the analyses in non-imputed and imputed data were statistically significant.

Cumulative effect of parental lifetime depression prenatally assessed on child problems at age 6 years

Finally, we graphically represented the contribution of each parents' prenatally assessed lifetime depression to child problems at age 6. Figure 1a shows that if both parents experienced depression, *children reported* more emotional problems. Importantly, the figure also clearly depicts that using *mother reports* of child problems gives different results: effects of maternal lifetime depression and of depression in two parents were larger than effects of fathers' depression only. Results for behavioral problems show a similar pattern (Figure 1b).

Table 3. Parental depressive symptoms at child age 3 years and child problems

	Child age 6 years				Child age 3 years			
	Child report (BPI)□		Mother report (CBCL)□		Father report (CBCL)□		Mother report (CBCL)□	
Depressive symptoms score, 3 years◊	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*
	Emotional problems score n = 3,131				*Emotional problems score n = 3,660*			
Father	0.02	-0.02;0.06	0.10#	0.07;0.14 ***	0.21#	0.18;0.25 ***	0.11#	0.07;0.14 ***
Mother	0.03	-0.01;0.06	0.18#	0.15;0.22 ***	0.14#	0.10;0.17 ***	0.20#	0.17;0.24 ***
	Behavioral problems score n = 3,121				*Behavioral problems score n = 3,660*			
Father	0.04	0.01;0.08 *	0.11#	0.07;0.15 ***	0.19#	0.16;0.23 ***	0.12#	0.08;0.15 ***
Mother	0.05	0.01;0.08 **	0.16#	0.13;0.20 ***	0.12#	0.08;0.15 ***	0.19#	0.15;0.22 ***

Note. Models adjusted for: parental age, income, marital status, maternal educational level, smoking during pregnancy, child national origin, child gender, child age, number of siblings and verbal abilities of the child. Shades indicate estimates where determinant and outcome were reported by the same informant.

□ BPI scores inverse transformed, CBCL scores square root transformed. All z-scores.

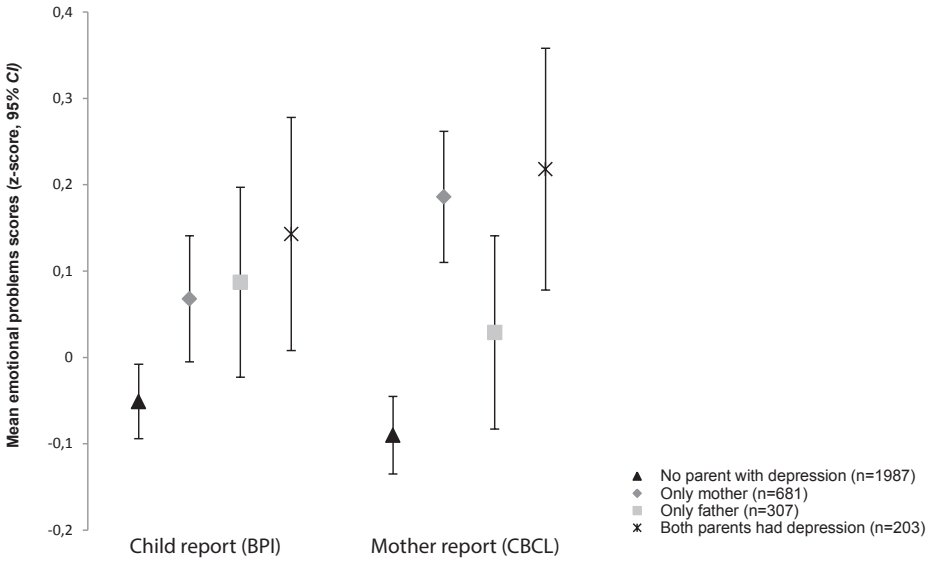
◊ Square root transformed, z-scores.

Significant difference in estimates for paternal and maternal depressive symptoms.

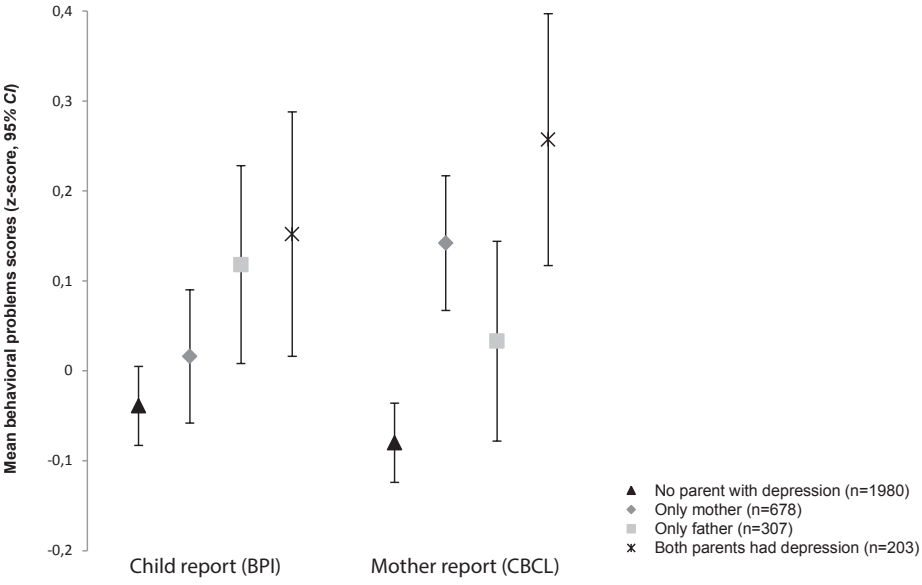
*$p < .05$, **$p < .01$, ***$p < .001$

Figure 1. Maternal versus paternal lifetime depression prenatally assessed and child problems at age 6 years

a. Emotional problems



b. Behavioral problems



Note. Estimates are means plus confidence intervals derived from ANCOVAs. Models adjusted for: parental age, income, marital status, maternal educational level, smoking during pregnancy, child national origin, child gender, child age, number of sibling, and verbal abilities of the child.

DISCUSSION

The present study showed that depression of mothers and of fathers is related to emotional and behavioral problems in three-to-six-year-old offspring. However, whether associations of paternal and maternal depression with child problems were similar, largely depended on the person reporting children's problems. When relying on children's self-reports of their problems, fathers' depression appeared to contribute to young children's problem behavior as much as mothers' depression. Whereas, if mothers or fathers reported on children's problems, the observed association of the respective parents' depression with child problems seemed inflated.

All informants agreed that both maternal and paternal depression impacted child behavior. Yet, results based on mother reported child problems suggest that fathers' depression has less impact on child well-being than mothers' depression. This is in line with our hypothesis and similar findings have been reported previously (Connell & Goodman, 2002; Malmberg & Flouri, 2011; Weitzman et al., 2011). However, when fathers reported child problems, we observed that the opposite should be concluded. Moreover, our results based on children's own reports showed again another picture: associations of paternal and maternal depression with child problems appeared very similar.

Different explanations for this inconsistency in findings are possible. Primarily it suggests that associations between one parents' depression and the same parents' report of child problems are inflated and reflect a bias. Importantly, this inflation was not specific for mother reported child problems: it was also present for father reports of child problems. Estimates were in some instances twice as large if parent reports were used to study the impact of a parent's own depression. This most probably resulted from shared-method variance, occurring if the same informant is used to report on determinant and outcome (Collishaw et al., 2009). Also, information-processing theories suggest that depressed individuals tend to interpret their experiences negatively (Foland-Ross & Gotlib, 2012), which can influence reports of depressed parents. Yet, parental depression was not measured concurrently with child problems at age six. Parents with a history of depression did not necessarily experience high levels of depressive symptoms while reporting child problems at age six.

The inconsistency between parent and child reports may also indicate that children are invalid reporters of their problems. The correlation between parental and child report were for example only modest. A child's developmental level, including short attention span, limited vocabulary and difficulties to report on complex constructs, is an inherent challenge to obtaining valid self-reports of young children (Luby et al., 2007; Measelle, Ablow, Cowan, & Cowan, 1998). However, if this were an issue, one would not expect to find very similar associations for paternal and maternal depression using child report, as well as very similar associations of paternal lifetime depression with child problems reported by the mother and the child him or herself. This suggests that fathers' depression may impact on young children's emotional and behavioral development as much as mothers' depression.

Each informant's report certainly has its own limitations. This is not only true in the present study, but has been found across studies examining diverse health related outcomes such as child

reported exposure to secondhand smoke and child and parent report of children's quality of life (Johnson-Kozlow et al., 2010; Robitail, Siméoni, Ravens-Sieberer, Bruil, & Auquier, 2007). Including children's information in addition to parent reports has shown to result in better estimates of an outcome (e.g. Arseneault et al., 2005; Johnson-Kozlow et al., 2010). Likewise, our results suggest that a more accurate picture of the effect of parental depression on child well-being can be obtained by comparing information from multiple sources. Moreover, the use of young child self-reports may provide researchers and practitioners with a surprisingly simple tool to circumvent the problem of shared informant bias if only parental reports are used.

In this study we examined outcomes at child age three years and child age six years. Since children are thought to become capable of reporting reliably about their emotions and behavior from about the age of about four years onward (Measelle et al., 1998), we used child report of problem behavior at child age six, but not at child age three years. Instead, we used father reports of child problems at child age three years to contrast results between informants. Mothers reported the outcome at both time points with the same instrument and results across the two time points were surprisingly consistent, even though children's emotional and behavioral problems are not necessarily stable across three years' time. Studies have reported that levels of behavioral problems tend to decrease in early childhood (Bongers, Koot, van der Ende, & Verhulst, 2003). For instance, aggressive behavior and temper tantrums of young children are transient in nature and decrease with age and developing self-regulation abilities. Emotional problems, on the other hand, seem to increase throughout young childhood (Bongers et al., 2003). This can be explained by the difficulties young children have communicating their emotions and with the covert nature of emotional problems. Despite potential changes in children's problems and changes in the presentation of these problems over time, the conclusions of the present study are similar across the two time points we examined.

Several explanations may underlie the negative effects of maternal and paternal depression on offspring well-being, including a genetically transmitted risk (Franić, Middeldorp, Dolan, Ligthart, & Boomsma, 2010; Singh et al., 2011; Sullivan, Neale, & Kendler, 2000), the impact of depression on parenting practices (Lovejoy, Graczyk, O'Hare, & Neuman, 2000), and hazardous influences of depression on family interactions (Low & Stocker, 2005). Despite associations of similar strength, dissimilar mechanisms may also account for the negative impact of maternal and paternal depression. Specific for maternal depression may be an intrauterine effect of depression during pregnancy on fetal neurodevelopment (Goodman & Gotlib, 1999). Children may also be more exposed to maternal depression, as mothers mostly spend more time with their children than fathers do. Potential explanations relating fathers' depression to child problems lie in the role fathers play in children's upbringing. Fathers are thought to encourage autonomy and self-control through play (i.e. rough-and-tumble play), which may prevent the development of aggressive behavior and anxiety for new situations (Paquette, 2004). Since fathers' depression affects parental play behaviors (Davis, Davis, Freed, & Clark, 2011; Wilson & Durbin, 2010), this may contribute to child problems. Additionally, depressive symptoms in men can elicit feelings like anger and irritability (Martin,

Neighbors, & Griffith, 2013), which may mediate the pathway between fathers' distress and child problems (Giallo, Cooklin, Wade, D'Esposito, & Nicholson, 2014).

The present study has considerable strengths, including the large sample size allowing for adjustment of a range of potential confounders, the inclusion of young children's self-reports of problems, and the prospective data assessment which reduced the likelihood of reversed causality. Nevertheless, several issues need to be discussed. First, obtaining a formal psychiatric diagnosis of lifetime depression in our large sample was not feasible. However, comparisons with CIDI-diagnoses in a subsample indicated that our assessment of lifetime depression had high sensitivity and specificity. Second, the non-response analysis indicated selective non-response. Participants included in the analyses had a higher socioeconomic status and were more often of Dutch national origin than persons who were lost to follow-up. Analyses were based on a selected group of families in which both mothers and fathers completed questionnaires at multiple time-points. It is possible that parents with severe depressive symptoms did not participate at baseline or dropped out of the study. This may have influenced the representativeness of our study. Moreover, participants with data on all time-points came from more advantageous backgrounds. Nevertheless, our results were similar in participants with data at either one of the time-points and participants with data at all time-points. This is in line with a results from a study by Wolke et al. (2009) who suggest that selective participation in cohort studies particularly constitutes a problem for estimating prevalence, but less likely biases the associations between variables. In addition, our results suggest that even in the more advantageous families, parental depression affects child well-being and that reporter bias may also influence studies with informants from a more advantageous background. Third, effect estimates were small, which is likely inherent to our predominantly healthy study sample. Also, the average level of postnatal depressive symptoms was lower in our study than in a Dutch norm sample, which may have contributed to a relatively small effect on child problems. Yet, small, but significant overall associations in a large non-clinical sample may well represent larger associations at the individual level, particularly in the presence of clinical depression. Finally, the majority of partners (99%) that filled out the questionnaire were the biological father of the children. In addition, more than 90% of the mothers lived together with a partner at outcome assessment. However, no data was available indicating whether mother's partner was indeed the biological father of the child at the time of outcome assessment. We adjusted analyses for living arrangement (i.e. did mother live together with a partner), but if the father filling out the questionnaire was not the partner living in the same family with the child, this may have resulted in an underestimation of the associations.

The current study demonstrates that both paternal and maternal depression are similarly associated with children's emotional and behavioral problems. Our findings also implicate that researchers and clinicians examining the impact of parental depression should thoroughly consider who is asked as an informant on child problems. Shared-method variance may have driven findings of previous studies which suggested that maternal depression affects child well-being more than paternal depression. To accurately estimate the effect of parental depression in research and in

clinical practice, information should ideally not be obtained from parents only. Rather, data from multiple sources should be considered, including young children's own valuable perspective on their problems.

REFERENCES

Ablow, J. C., & Measelle, J. R. (2003). *Manual for the Berkeley Puppet Interview: Symptomatology, social, and academic modules (bpi 1.0)*. MacArthur foundation Research Network Psychopathology and Development (David J. Kupfer, Chair): The MacArthur Working Group on Outcome Assessment

Ablow, J. C., Measelle, J. R., Kraemer, H. C., Harrington, R., Luby, J., Smider, N., Dierker, L., Clark, V., Dubicka, B., Heffelfinger, A. M. Y., Essex, M. J., & Kupfer, D. J. (1999). The MacArthur Three-City Outcome Study: Evaluating multi-informant measures of young children's symptomatology. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(12), 1580-1590.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth & Families.

Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffitt, T. (2005). Psychometric evaluation of 5- and 7-year-old children's self-reports of conduct problems. *Journal of Abnormal Child Psychology, 33*(5), 537-550.

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2003). The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology, 112*(2), 179-192.

Cabrera, N., Tamis-LeMonda, C. S., Bradley, R. H., Hofferth, S., & Lamb, M. E. (2000). Fatherhood in the twenty-first century. *Child Development, 71*(1), 127-136.

Collishaw, S., Goodman, R., Ford, T., Rabe-Hesketh, S., & Pickles, A. (2009). How far are associations between child, family and community factors and child psychopathology informant-specific and informant-general? *Journal of Child Psychology and Psychiatry, 50*(5), 571-580.

Compas, B., Phares, V., Banez, G., & Howell, D. (1991). Correlates of internalizing and externalizing behavior problems: Perceived competence causal attributions, and parental symptoms. *Journal of Abnormal Child Psychology, 19*(2), 197-218.

Connell, A. M., & Goodman, S. H. (2002). The association between psychopathology in fathers versus mothers and children's internalizing and externalizing behavior problems: A meta-analysis. *Psychological bulletin, 128*(5), 746.

Davis, R. N., Davis, M. M., Freed, G. L., & Clark, S. J. (2011). Fathers' depression related to positive and negative parenting behaviors with 1-year-old children. *Pediatrics, 127*(4), 612-618.

De Beurs, E. (2004). Brief Symptom Inventory. *Leiden, Netherlands: Pits Publishers*.

Foland-Ross, L. C., & Gotlib, I. H. (2012). Cognitive and neural aspects of information processing in major depressive disorder: An integrative perspective. *Frontiers in psychology, 3*.

Franić, S., Middeldorp, C. M., Dolan, C. V., Ligthart, L., & Boomsma, D. I. (2010). Childhood and adolescent anxiety and depression: Beyond heritability. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(8), 820-829.

Gere, M. K., Hagen, K. A., Villabø, M. A., Arnberg, K., Neumer, S.-P., & Torgersen, S. (2013). Fathers' mental health as a protective factor in the relationship between maternal and child depressive symptoms. *Depression and Anxiety, 30*(1), 31-38.

Giallo, R., Cooklin, A., Wade, C., D'Esposito, F., & Nicholson, J. M. (2014). Maternal postnatal mental health and later emotional–behavioural development of children: The mediating role of parenting behaviour. *Child: Care, Health and Development, 40*(3), 327-336.

Goldwasser, M. A., & Fitzmaurice, G. M. (2001). Multivariate linear regression analysis of childhood psychopathology using multiple informant data. *International Journal of Methods in Psychiatric Research, 10*(1), 1-10.

Goodman, S. H., & Gotlib, I. H. (1999). Risk for psychopathology in the children of depressed mothers: A developmental model for understanding mechanisms of transmission. *Psychological review, 106*(3), 458.

Jaddoe, V. W., van Duijn, C. M., Franco, O. H., van der Heijden, A. J., van IJzendoorn, M. H., de Jongste, J. C., van der Lugt, A., Mackenbach, J. P., Moll, H. A., Raat, H., Rivadeneira, F., Steegers, E. A., Tiemeier, H., Uitterlinden, A. G., Verhulst, F. C., & Hofman, A. (2012). The Generation R Study: Design and cohort update 2012. *European Journal of Epidemiology, 27*(9), 739-756.

Johnson-Kozlow, M., Wahlgren, D. R., Hovell, M. F., Flores, D. M., Liles, S., Hofstetter, C. R., Zellner, J., & Zakarian, J. M. (2010). Adolescents validly report their exposure to secondhand smoke. *Journal of Clinical Epidemiology, 63*(8), 914-919.

Kvalevaag, A. L., Ramchandani, P. G., Hove, O., Assmus, J., Eberhard-Gran, M., & Biringer, E. (2013). Paternal mental health and socioemotional and behavioral development in their children. *Pediatrics, 131*(2), e463-e469.

LaGrange, B., & Cole, D. A. (2008). An expansion of the trait-state-occasion model: Accounting for shared method variance. *Structural Equation Modeling, 15*(2), 241-271.

Lovejoy, M. C., Graczyk, P. A., O'Hare, E., & Neuman, G. (2000). Maternal depression and parenting behavior: A meta-analytic review. *Clinical Psychology Review, 20*(5), 561-592.

Low, S. M., & Stocker, C. (2005). Family functioning and children's adjustment: Associations among parents' depressed mood, marital hostility, parent-child hostility, and children's adjustment. *Journal of Family Psychology, 19*(3), 394-403.

Luby, J. L., Belden, A., Sullivan, J., & Spitznagel, E. (2007). Preschoolers' contribution to their diagnosis of depression and anxiety: Uses and limitations of young child self-report of symptoms. *Child Psychiatry and Human Development, 38*(4), 321-338.

Malmberg, L. E., & Flouri, E. (2011). The comparison and interdependence of maternal and paternal influences on young children's behavior and resilience. *Journal of Clinical Child & Adolescent Psychology, 40*(3), 434-444.

Martin, L. A., Neighbors, H. W., & Griffith, D. M. (2013). The experience of symptoms of depression in men vs women: Analysis of the national comorbidity survey replication. *JAMA Psychiatry, 70*(10), 1100-1106.

Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing young children's views of their academic, social, and emotional lives: An evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development, 69*(6), 1556-1576.

Papp, L. (2012). Longitudinal associations between parental and children's depressive symptoms in the context of interparental relationship functioning. *Journal of Child and Family Studies, 21*(2), 199-207.

Paquette, D. (2004). Theorizing the father-child relationship: Mechanisms and developmental outcomes. *Human Development, 47*(4), 193-219.

Pearson, R. M., Evans, J., Kounali, D., & et al. (2013). Maternal depression during pregnancy and the postnatal period: Risks and possible mechanisms for offspring depression at age 18 years. *JAMA Psychiatry, 70*(12), 1312-1319.

Ramchandani, P., O'Connor, T., Evans, J., Heron, J., Murray, L., & Stein, A. (2008a). The effects of pre- and postnatal depression in fathers: A natural experiment comparing the effects of exposure to depression on offspring. *Journal of Child Psychology and Psychiatry, 49*(10), 1069-1078.

Ramchandani, P., & Psychogiou, L. (2009). Paternal psychiatric disorders and children's psychosocial development. *The Lancet, 374*(9690), 646-653.

Ramchandani, P., Stein, A., Evans, J., & O'Connor, T. G. (2005). Paternal depression in the postnatal period and child development: A prospective population study. *The Lancet, 365*(9478), 2201-2205.

Ramchandani, P., Stein, A., O'Connor, T. G., Heron, J. O. N., Murray, L., & Evans, J. (2008b). Depression in men in the postnatal period and later child psychopathology: A population cohort study. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(4), 390-398.

Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin, 112*(3), 485-499.

Ringoot, A. P., Jansen, P. W., Steenweg-de Graaff, J., Measelle, J. R., van der Ende, J., Raat, H., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2013). Young children's self-reported emotional, behavioral, and peer problems: The Berkeley Puppet Interview. *Psychological Assessment, .25*(4)

Robitail, S., Siméoni, M.-C., Ravens-Sieberer, U., Bruil, J., & Auquier, P. (2007). Children proxies' quality-of-life agreement depended on the country using the European Kidscreen-52 questionnaire. *Journal of Clinical Epidemiology, 60*(5), 469.e461-469.e413.

Singh, A. L., D'Onofrio, B. M., Slutske, W. S., Turkheimer, E., Emery, R. E., Harden, K. P., Heath, A. C., Madden, P. A., Statham, D. J., & Martin, N. G. (2011). Parental depression and offspring psychopathology: A children of twins study. *Psychological Medicine, 41*(7), 1385-1395.

Statistics Netherlands. (2004). Allochtonen in Nederland 2004 [Foreigners in the Netherlands 2004]. Voorburg/ Heerlen.

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry, 157*(10), 1552-1562.

Tharner, A., Luijk, M. P. C. M., Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2012). Maternal lifetime history of depression and depressive symptoms in the prenatal and early postnatal period do not predict infant–mother attachment quality in a large, population-based dutch cohort study. *Attachment & Human Development, 14*(1), 63-81.

Van Bon, W. H. J., & Hoekstra, J. G. (1982). *Taaltests voor kinderen*. Lisse: Swets & Zeitlinger.

Velders, F. P., Dieleman, G., Henrichs, J., Jaddoe, V. W., Hofman, A., Verhulst, F. C., Hudziak, J. J., & Tiemeier, H. (2011). Prenatal and postnatal psychological symptoms of parents and family functioning: The impact on child emotional and behavioural problems. *European Child & Adolescent Psychiatry, 20*(7), 341-350.

Weissman, M., Wickramaratne, P., Nomura, Y., Warner, V., Pilowsky, D., & Verdeli, H. (2006). Offspring of depressed parents: 20 years later. *American Journal of Psychiatry, 163*(6), 1001-1008.

Weitzman, M., Rosenthal, D. G., & Liu, Y.-H. (2011). Paternal depressive symptoms and child behavioral or emotional problems in the united states. *Pediatrics, 128*(6), 1126-1134.

WHO. (1990). *Composite international diagnostic interview (cidi): A) cidi-interview (version 1.0), b) cidi-user manual, c) cidi-training manual, d) cidi-computer programs* Geneva: World Health Organization.

Wilson, S., & Durbin, C. E. (2010). Effects of paternal depression on fathers' parenting behaviors: A meta-analytic review. *Clinical Psychology Review, 30*(2), 167-180.

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., & Lamberts, K. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *British Journal of Psychiatry, 195*(3), 249-256.

SUPPLEMENTARY MATERIAL

Figure S1. Participant flow starting from prenatal assessment, subsample one

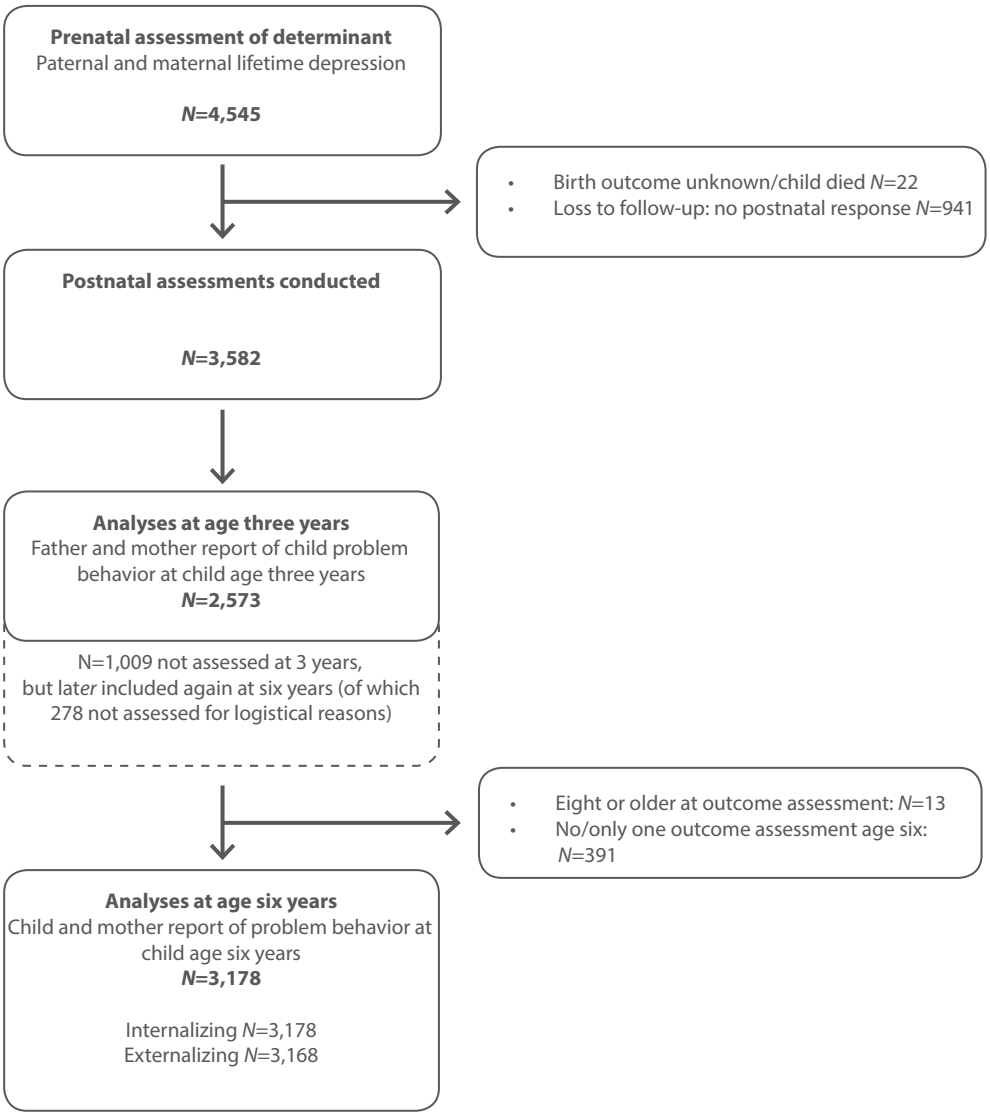


Figure S2. Participant flow starting from postnatal parental depression data, subsample two

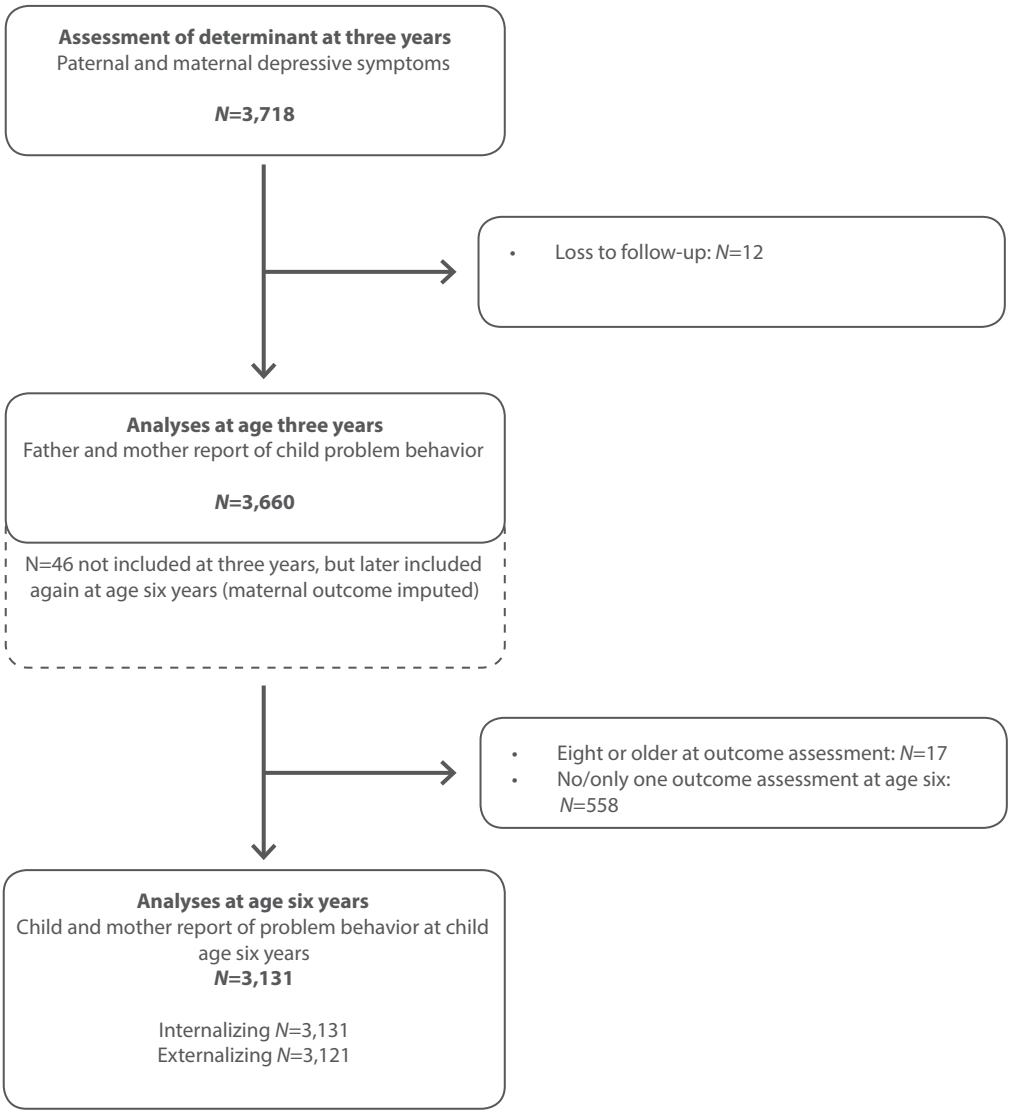


Table S1. Composite scores of child problems and formal tests of informant differences in mother-child and mother-father reports of child problems.

	Child age 6 years						Child age 3 years					
	Emotional problems score□			Behavioral problems score□			Emotional problems score□			Behavioral problems score□		
	n	*B (95% CI)*	Diff. child/ mother	*n*	*B (95% CI)*	Diff. child/ mother	*n*	*B (95% CI)*	Diff. child/ mother	*N*	*B (95% CI)*	Diff. child/ mother
Lifetime depression, prenatally	3,178			3,168			2,573			2,573		
Father		0.11 (0.04;0.18)**	C=M		0.15 (0.08;0.22)**	C=M		0.19 (0.10;0.28)**	F=M		0.19 (0.10;0.29)**	F=M
Mother		0.19 (0.13;25)**	C<M**		0.19 (0.13;0.25)**	C<M**		0.20 (0.13;0.28)**	F<M*		0.19 (0.11;0.27)**	F=M
Depressive symptoms◊ (age 3 years)	3,131			3,121			3,660			3,660		
Father		0.08 (0.06;0.11)**	C<M**		0.09 (0.06;0.11)**	C<M**		0.18 (0.15;0.21)**	F>M**		0.17 (0.14;0.19)**	F>M**
Mother		0.13 (0.10;0.15)**	C<M**		0.12 (0.09;0.15)**	C<M**		0.19 (0.17;0.22)**	F<M**		0.17 (0.14;0.19)**	F<M**

Note. 'C=M' or 'F=M' non-significant interaction: association of depression with child problems does not depend on informant; 'C<M', 'F>M' or 'F<M' statistically significant interaction: association of depression with child problems depends on informant.

□BPI scores inverse transformed, CBCL scores square root transformed. Both z-scores.

◊ Square root transformed, z-scores.

*p-value <.05; **p-value <.01

Table S2. Parental prenatal lifetime depression and child problems

	Child age 6 years (n = 2,091)				Child age 3 years (n = 2,079)‡			
	Child report (BPI)□		Mother report (CBCL)□		Father report (CBCL)□		Mother report (CBCL)□	
Lifetime depression, prenatally (yes/no)	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*
	Emotional problems score				*Emotional problems score*			
Father	0.16	0.05;0.28**	0.12	0.01;0.23*	0.28	0.16;0.39***	0.19	0.07;0.30**
Mother	0.10	0.01;0.19*	0.25	0.16;0.35***	0.18	0.08;0.27***	0.26	0.17;0.35***
	Behavioral problems score				*Behavioral problems score*			
Father	0.14	0.03;0.26*	0.16	0.04;0.27**	0.24	0.12;0.35***	0.21	0.10;0.32***
Mother	0.05	-0.05;0.14	0.24	0.15;0.33***	0.16	0.07;0.26**	0.23	0.13;0.32***

Note. Models adjusted for: parental age, income, marital status, maternal educational level, smoking during pregnancy, child national origin, child gender, child age, number of siblings and verbal abilities of the child.
Shades indicate estimates where determinant and outcome were reported by the same informant.
□BPI scores inverse transformed, CBCL scores square root transformed. All z-scores.
‡ In these analyses only 2079 persons could be included, as a few participants had all determinants and complete outcome at age 6, but had neither maternal or paternal CBCL data at age 3 years.
*p <.05, **p <.01, ***p <.001

Table S3. Parental depressive symptoms at child age 3 years and child problems

	Child age 6 years (n = 2,091)				Child age 3 years (n = 2,079)‡			
	Child report (BPI)□		Mother report (CBCL)□		Father report (CBCL)□		Mother report (CBCL)□	
Depressive symptoms score, 3 years◊	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*	*B*	*95% CI*
	Emotional problems score				*Emotional problems score*			
Father	0.03	-0.01;0.07	0.10#	0.05;0.14***	0.19#	0.15;0.24***	0.10#	0.06;0.14***
Mother	0.05	0.01;0.10*	0.19#	0.15;0.23***	0.11#	0.07;0.16***	0.19#	0.15;0.24***
	Behavioral problems score				*Behavioral problems score*			
Father	0.05	0.01;0.09*	0.12#	0.07;0.16***	0.17	0.13;0.21***	0.12#	0.08;0.16***
Mother	0.07	0.02;0.11**	0.19#	0.14;0.23***	0.11	0.06;0.15***	0.20#	0.16;0.25***

Note. Models adjusted for: parental age, income, marital status, maternal educational level, smoking during pregnancy, child national origin, child gender, child age, number of siblings and verbal abilities of the child.
Shades indicate estimates where determinant and outcome were reported by the same informant.
□ BPI scores inverse transformed, CBCL scores square root transformed. All z-scores.
◊ Square root transformed, z-scores.
Significant difference in estimates for paternal and maternal depressive symptoms
‡In these analyses only 2079 persons could be included, as a few participants had all determinants and complete outcome at age 6, but had neither maternal or paternal CBCL data at age 3 years.
*p <.05, **p <.01, ***p <.001

chapter 6

Exploring the relation of harsh parental discipline with child emotional and behavioral problems by using multiple informants: The Generation R Study

Mackenbach, J.D., Ringoot, A.P., Ende, van der, J., Verhulst, F.C., Jaddoe, V.W.V., Hofman, A., Jansen, P.W., & Tiemeier, H. *(2014)*

PlosOne. 9(8), e104793.

ABSTRACT

Parental harsh discipline, like corporal punishment, has consistently been associated with adverse mental health outcomes in children. It remains a challenge to accurately assess the consequences of harsh discipline, as researchers and clinicians generally rely on parent report of young children's problem behaviors. If parents rate their parenting styles and their child's behavior this may bias results. The use of child self-report of problem behaviors is not common but may provide extra information about the relation of harsh parental discipline and problem behavior. We examined the independent contribution of young children's self-report above parental report of emotional and behavioral problems in a study of maternal and paternal harsh discipline in a birth cohort. Maternal and paternal harsh discipline were associated with both parent reported behavioral and parent reported emotional problems, but only with child reported behavioral problems. Associations were not explained by pre-existing behavioral problems at age three. Importantly, the association with child reported outcomes was independent from parent reported problem behavior. These results suggest that young children's self-reports of behavioral problems provide unique information on the effects of harsh parental discipline. Inclusion of child self-reports can therefore help estimate the effects of harsh parental discipline more accurately.

INTRODUCTION

Parenting practices play a fundamental role in children's emotional and behavioral development. Corporal disciplining practices have consistently been associated with adverse mental health outcomes, such as poor school achievements, behavioral problems, lowered self-esteem and delinquent behaviors (Berger, 2005; Larzelere, 2000; Sidebotham & Golding, 2001; Teicher, Samson, Polcari, & McGreenery, 2006). Milder forms of negative parental disciplining strategies -like harsh discipline- have also been studied repeatedly. Harsh discipline is characterized by parental attempts to control a child using verbal violence (shouting) or physical forms of punishment (pinching or hitting; Chang, Schwartz, Dodge, & McBride-Chang, 2003). These forms of parental discipline have been associated not only with child behavioral problems, in line with a cycle of violence hypothesis (Avakame, 1998), but also with child emotional problems (Chang et al., 2003; Taylor, Manganello, Lee, & Rice, 2010; Teicher et al., 2006; Vostanis et al., 2006). The effects of these milder forms of harsh discipline may be less pronounced, yet are important since the prevalence of these forms of parental discipline is high. In a recent study using data from the present cohort we demonstrated that no less than 77% of mothers and 67% of fathers shouted at their child at least once in the last two weeks, in addition the number of parents threatening to slap (20-24%) or angrily pinching the child's arm (15%) was also considerable (Jansen et al., 2012). Given the high prevalence and the known burden for children, it is important to examine the consequences of these milder forms of harsh parental discipline accurately.

As child behavior problems like aggressive or oppositional behaviors may lead to higher levels of harsh discipline by parents (Anderson, Lytton, & Romney, 1986), it is important to study the effects of harsh parental discipline on child problem behaviors prospectively. A number of longitudinal studies have affirmed that, after controlling for baseline emotional and behavioral problems, children exposed to less extreme forms of parental harsh discipline have an increased risk of behavioral problems and psychiatric disorders later in life (Larzelere, 2000; Taylor et al., 2010).

Despite a large body of evidence, the existing literature on emotional and behavioral consequences of mild harsh discipline suffers limitations. Most studies relied on parental and often only on maternal report of child behavioral problems (Lansford et al., 2005; Larzelere, 2000; McLeod, Kruttschnitt, & Dornfeld, 1994; Prinzie, Onghena, & Hellinckx, 2006; Taylor et al., 2010). Relying on one informant for both the determinant and the outcome is problematic, as parents who rate their own parenting styles as 'harsh' may also perceive their child's behavior differently than parents that do not use harsh discipline (Anderson et al., 1986; Chang et al., 2003; Fox, Platz, & Bentley, 1994; Reid, Kavanagh, & Baldwin, 1987). This problem can be avoided if the informant reporting on the consequences of harsh discipline differs from the informant reporting on harsh discipline. Including multiple reporters may generate additional evidence regarding the consequences of harsh discipline. It has become widely accepted that young children may be a valuable source of information (Kraemer et al., 2003), as they can provide unique insights into their own behaviors (Arsenault et al., 2003). Indeed, self-report on the consequences of harsh discipline has proven to generate valuable results in adolescents (Bolger & Patterson, 2001; Lansford et al., 2005; Lansford,

Deater-Deckard, Dodge, Bates, & Pettit, 2004). Yet, few studies on the consequences of parental harsh discipline have used young children's self-reports.

In the present study we examined the consequences of both maternal and paternal harsh discipline on parent reported and young children's self-reported emotional and behavioral problems. Specifically, we investigated whether any effect observed using child report was independent of parent reported problems. We hypothesized that child self-report of problem behavior would strengthen the evidence of an association between harsh discipline and parent reported problem behavior by contributing unique information.

METHODS

Ethics statement

The study was conducted in accordance with the guidelines proposed in the World Medical Association Declaration of Helsinki and has been approved by the Medical Ethical Committee of the Erasmus Medical Center in Rotterdam, the Netherlands (MEC 198.782/2001/31). Full written informed consent for the postnatal phase was obtained from parents for both parental and child data.

Study design and population

This study was embedded in the Generation R Study, a prospective population-based cohort from fetal life onwards. The design and data collection methods have been extensively described elsewhere (Jaddoe et al., 2012). Briefly, all pregnant women residing in Rotterdam, with an expected delivery date between April 2002 and January 2006, were eligible for participation in Generation R. For this study, we considered participants with full postnatal written consent ($N = 7,295$) eligible. A questionnaire including parental discipline at age 3 years was returned by 4,733 mothers and constituted the baseline. Of those, 718 children had missing data on child self-reported emotional and behavioral problems (BPI) at age 6 years, yielding a sample size of 4,015 (follow-up response: 85%) for analyses with maternal harsh discipline and *child reported problems*. The sample size for analyses with maternal harsh discipline and *parent reported problems* was $n = 3,764$. A flow chart is provided in supplementary material (Figure S1).

Measures

Harsh Discipline

Information about parental harsh discipline was obtained by postal questionnaires when the children were 3 years old. We assessed various types of disciplining practices by ten items that were based on the Parent-Child Conflict Tactics Scale (Straus, Hamby, Moore, Runyan, 1998). In a previous study in the same cohort, a harsh discipline scale was confirmed using factor analysis (Jansen et al., 2012). This resulted in a scale consisting of 6 items, representing constructs of psychological aggression and (mild) physical assault: "In the past week/month, I angrily pinched my child's arm",

"I shouted, yelled or screamed angrily at my child", "I scolded at my child", "I threatened to slap, spank or hit my child but did not actually do it", "I called my child dumb or lazy or some other name like that" and "I shook my child". Items were scored on a scale from 0 to 2. In line with this previous study (Jansen et al., 2012), we calculated separate maternal and paternal harsh discipline scores by summing these 6 items. This yielded a score ranging from 0 to 12, with higher scores reflecting higher severity of harsh discipline.

Emotional and behavioral problems

Children were invited to our research centre in Rotterdam at the age of 6 years. During this visit, the Berkeley Puppet Interview (BPI) was used to assess emotional and behavioral problems as perceived by the child him/herself as described previously (Ringoot et al., 2013). The BPI is a semi-structured interactive interview technique to obtain self-reports of young children. During the interview, two identical dog hand puppets were introduced to the child and invited the child to engage in a conversation. The puppets made opposing statements about themselves. For example, one puppet said that he was a sad kid, while the other puppet said that he was not a sad kid. Subsequently, the puppets asked children to indicate which statement described themselves best. In this study, we used internalizing (emotional problems) and externalizing (behavioral problems) scales. The Internalizing scale score (20 items) was computed as the sum of the item scores in three scales: Depression, Separation Anxiety and Overanxious. The Externalizing scale score (21 items) was computed as the sum of the item scores in three scales: Oppositional Defiant, Overt Hostility and Conduct Problems. Higher scores on the BPI scales indicate more problems. The psychometric properties of the BPI emotional and behavioral scales in the present study have been described elsewhere (Luby, Belden, Sullivan, & Spitznagel, 2007; Ringoot et al., 2013).

Parent-reported child emotional and behavioral problems were assessed with the Dutch version of the Child Behavior Checklist (CBCL/1,5-5), a 99-item questionnaire that was mailed prior to the visit to the research centre (Achenbach & Rescorla, 2001). One of the parents, usually the mother, completed the CBCL/1,5-5 just before the visit to the research centre (92% of the questionnaires were completed by the mothers, 8% by other (primary) caregivers). The internalizing (emotional problems) and externalizing (behavioral problems) broadband scales were used in the present study. The Internalizing scale score (36 items) is the sum of the item scores of four scales: Emotionally Reactive, Anxious/Depressed, Somatic Complaints, and Withdrawn. The Externalizing scale score (24 items) is the sum of the item scores of the Attention Problems and Aggressive Behavior scales. Higher scores on the CBCL scales indicated more problems. Good reliability and validity has been reported for the CBCL/1,5-5 (Tick, Ende, van der, Koot, & Verhulst, 2007).

Assessing child problems at age 6 years was considered appropriate, as both the BPI and the CBCL are valid tools to assess child emotions and behaviors at this age (Ringoot et al., 2013; Rescorla et al., 2007). Pre-existing child internalizing and externalizing problems were reported by both mother and father using the CBCL/1,5-5 when children were 3 years old. This provided a 3 year difference, during transition from preschool to school-age, between determinant and outcome.

Covariates

Potential confounders were selected based on prior studies (Day, Peterson, & Mccracken, 1998; Taylor et al., 2010; Vostanis et al., 2006). Information on gender, date of birth, marital status of the parents, smoking during pregnancy and age of the parents at intake was obtained from midwifery and hospital registries. Information on ethnicity, number of children in the household, educational level of the parents and household income was obtained by questionnaires at age six years. The child's ethnicity was classified by the countries of birth of the parents, according to the Dutch standard classification criteria of Statistics Netherlands (2004), and was categorized into Dutch, European and Non-Western background (e.g. Turkish, Moroccan, Indonesian, Cape Verdean, Surinamese and Antillean). Educational level of the parents was defined as low (at most lower vocational training), medium low (at most intermediate vocational training), medium high (higher vocational training) and high (university degree). Family household income was divided into two categories: below 2,000 Euros per month which corresponds with below modal income, and 2,000 Euros per month and above. Marital status of parents was defined as either being married/living together or as having no partner.

To assess global parental psychopathology, a selection of 21 items from the Brief Symptom Inventory (BSI; Boulet & Boss, 1991) was administered to both mothers and fathers when the child was three years old.

Family functioning was measured with the 12-item General Functioning scale of the McMasters Family Assessment Device (FAD; Miller, Epstein, Bishop, Keitner, 1985). In this validated self-report questionnaire, parents (in 82% of the cases this was the mother) rated family functioning and family stress on a 4-point scale.

Statistical analyses

We first conducted descriptive analyses of the population. Next, correlational analysis of harsh parenting, emotional (internalizing) and behavioral (externalizing) problems and parental psychopathology was performed.

The relation between harsh discipline at age three years and emotional and behavioral problems at age six years was examined with linear regression analyses. To satisfy the assumption of normality, maternal and paternal harsh discipline scores were square root transformed to achieve a normal distribution. Similarly, BPI and CBCL scale scores were transformed using the natural logarithm and the square root respectively, and z-scores were calculated to be able to compare the emotional and behavioral problems with each other.

We studied the effects of maternal and paternal harsh discipline separately. Similarly, we studied parent and child self-reports of emotional and behavioral problems as separate outcomes.

In model 1, unadjusted linear regression analyses of harsh discipline with child emotional and behavioral problems were performed. In model 2, we adjusted for sociodemographic characteristics (child gender, age and ethnicity, number of children in the household, household income, marital status, smoking during pregnancy, maternal and paternal educational level), maternal and paternal psychopathology score, and family functioning. Covariates were included in the second model if they changed the effect estimates of the unadjusted relation between harsh discipline and

emotional and behavioral problems by more than 5%. To adjust for pre-existing emotional and behavioral problems, in model 3 we additionally accounted for emotional (if emotional problems were the outcome) or behavioral problems (if behavioral problems were the outcome) assessed at age three years. If the association between harsh discipline and problem behavior is independent of baseline problem behavior, this would strengthen the assumption concerning the temporality of the associations.

Analyses were adjusted for maternal characteristics (maternal education, maternal psychopathology and the maternal report of pre-existing child emotional/behavioral problems) unless paternal harsh discipline was the independent variable; in this case we adjusted for the respective paternal characteristics.

Next, we additionally adjusted the analyses of the child self-report problem behavior (model 3) for parent reported emotional and behavioral problems (model 4). The aim of this analysis was to examine whether harsh discipline could predict child self-reported emotional and behavioral problems, over and above parent report. If the association is independent of parent report, this suggests children can provide unique outcome information in this study of parental harsh discipline.

To test the influence of effect modifiers, we specified interaction terms for harsh discipline with child gender, child ethnicity and a mutual interaction term between maternal harsh discipline and paternal harsh discipline on the risk of emotional and behavioral problems. None of the interaction terms was statistically significant.

Missing values on the covariates were estimated using multiple imputation techniques and were based on available information on determinants, outcome and covariates of this study. The presented results are based on pooled estimates of ten imputed datasets.

Analyses were conducted in the number of children with data available for the outcome of interest (for example parent reported emotional problems). As we did not impute the outcome variables the number of children per analysis differed from 3,047 to 4,015. We repeated all analyses in participants with complete data (n = 3,047). Linear regression analyses were performed using the SPSS version 18.0 (SPSS Inc., Chicago, IL).

Baseline nonresponse and loss-to-follow up analysis

In total, 4,733 mothers completed the questionnaire on harsh discipline at baseline. Mothers (n = 2,562) who did not complete this questionnaire were on average younger (28.6 years versus 31.5 years, $F(2, n = 7{,}295) = 11.8$, $p < .001$), and were more likely to have continued smoking during pregnancy (20.8% versus 12.0%, $\chi^2(2, n = 7{,}295) = 501.0$, $p < .001$), to have a family income below modal (32.7% versus 15.4%, $\chi^2(1, n = 7{,}295) = 238.5$ $p < .001$) and to have no partner (21.4% versus 8.2%, $\chi^2(1, n = 7{,}295) = 231.8$, $p < .001$) than mothers who completed the questionnaire.

At follow-up, when the child was between five and eight years old, 4,015 children (85%) of the 4,733 mothers who returned the questionnaire at age three, completed the Berkeley Puppet Interview. We compared these families with families of children who did not complete the BPI (n = 718). Children without a BPI assessment were more likely to be of non-Dutch origin (38.3% versus 32.2%, $\chi^2(2, n = 4{,}733) = 11.8$, $p = .003$), but did not differ from their peers who completed a BPI assessment in terms of maternal harsh discipline score (2.2 versus 2.2, $F = 0.7$, $p = .63$), behavioral

problems at age three (5.4 versus 5.0, $F = 10.2$, $p = .08$), parent reported behavioral problems at age six (7.5 versus 6.9, $F = 12.4$, $p = .13$) or family income (14.3% versus 15.6%, below modal, $\chi^2(1, n = 4{,}733) = 0.8$, $p = .40$).

Table 1. Characteristics of the study population $N = 4{,}015$

Child & family characteristics	Percentages or means (*SD*)
Gender (% boys)	49.9%
Age at BPI[¥] measurement in years	6.03 (0.35)
Child ethnicity	
Dutch	67.4%
European	8.2%
Non-Western	24.4%
Number of children in the household	2.50 (1.72)
Age mother at intake	31.69 (4.54)
Age partner at intake	34.09 (5.28)
Harsh discipline by mother	2.18 (1.95)
Harsh discipline by father	1.82 (1.81)
Household income (% above modal)	68.7%
Marital status (% with partner)	92.8%
Highest educational level of parents	
Low	4.4 %
Medium low	19.9 %
Medium high	26.5 %
High	49.3 %
Smoking during pregnancy	
Never	79.3 %
Until pregnancy was known	9.1 %
Continued during pregnancy	11.6 %
Family functioning score	1.49 (0.41)
Psychopathology of mother score	3.43 (5.62)
Psychopathology of father score	2.71 (4.45)
Parent reported CBCL scores	
Emotional problems score[□] at age 3	5.15 (4.39)
Behavioral problems score[□] at age 3	8.76 (5.71)
Emotional problems score[□] at age 6	5.44 (5.28)
Behavioral problems score[□] at age 6	6.94 (5.28)
Child self-reported BPI scores	
Emotional problems score[‡] at age 6	58.16 (12.10)
Behavioral problems score[‡] at age 6	51.92 (10.60)

[□] measured by the CBCL (parent report), ‡ measured by the BPI (child self-report)

RESULTS

Characteristics of the study sample are presented in Table 1. Children had a mean age of 3.1 years at baseline and a mean age of 6.1 years at follow-up. Sixty-seven percent of the children were of Dutch origin, 8.2% had a European and 24.4% a Non-western background.

Table 2 shows the Pearson correlation coefficients between harsh discipline, the different emotional and behavioral problem scales, and parental psychopathology. Parent reports of emotional and behavioral problems were highly correlated (r at age 3 years = .61, $p < .001$, r at age 6 years = .66, $p < .001$), whereas child reported emotional and behavioral problems were less strongly correlated ($r = .30$, $p < .001$). Parent and child reports of behavioral problems were more strongly correlated ($r = .18$, $p < .001$) than emotional problems ($r = .10$, $p < .001$). Maternal and paternal harsh discipline was correlated with all emotional and behavioral scales, with the exception that there was no correlation between paternal harsh discipline and child reported emotional problems.

Table 2. Correlational analysis of harsh discipline, emotional and behavioral problems and parental psychopathology

	M	*SD*	*1.*	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*
1. CBCL□ emotional sumscore age 3‡	5.09	3.86	1								
2. CBCL□ behavioral sumscore age 3‡	8.79	5.27	.61**	1							
3. CBCL□ emotional sumscore age 6	5.41	5.24	.49**	.40**	1						
4. CBCL□ behavioral sumscore age 6	6.92	6.05	.35**	.54**	.66**	1					
5. BPI‡ emotional sumscore age 6	57.95	12.08	.08**	.07**	.10**	.08**	1				
6. BPI‡ behavioral sumscore age 6	51.84	10.56	.03*	.13**	.07**	.18**	.30**	1			
7. Maternal harsh discipline	2.16	1.91	.22**	.35**	.19**	.26**	.05*	.09*	1		
8. Paternal harsh discipline	1.80	1.81	.20**	.37**	.12**	.23**	.02	.11*	.38*	1	
9. Maternal psychopathology	3.33	5.39	.32**	.30**	.28**	.26**	.05*	.04*	.30*	.12*	1
10. Paternal psychopathology	2.68	4.39	.30**	.28**	.16**	.17**	.06*	.06*	.12*	.26*	.28*

□ Child Behavior Checklist (parent report), ◊ Mean of mother and father score on the internalizing/externalizing scale,
‡ Berkeley Puppet Interview (child self-report) *p-value < .05, **p-value < .001

Table 3 shows the results of the linear regression analyses with *behavioral problems* as outcome. First, we assessed the relation between harsh discipline and *parent reported behavioral problems* (CBCL). Adjustment for sociodemographic covariates and family characteristics attenuated the effect of harsh discipline on behavioral problems. Additional adjustment for baseline behavioral problems at age 3 further attenuated the effect estimates, but the relation between *maternal* harsh discipline and parent reported behavioral problems remained (model 3: *B* = 0.05, *95% CI*: 0.02, 0.09). Analyses of the relation between *paternal* harsh discipline and behavioral problems yielded similar results (model 3: *B* = 0.08, *95% CI*: 0.04, 0.13).

Analyses with *child self-reported behavioral problems* (BPI) showed that higher levels of maternal harsh discipline were associated with higher levels of child reported behavioral problems. Although effect sizes were somewhat smaller than those for parent reported problems, the overall pattern for child reported behavioral problems across the three models was very similar to the effect observed if based on parent report. Even after adjustment for all covariates, *maternal* and *paternal* harsh discipline were associated with a higher score on child self-reported behavioral problems (model 3: *B* for maternal harsh discipline = 0.07, *95% CI*: 0.03, 0.11; *B* for paternal harsh discipline = 0.09, *95% CI*: 0.04, 0.13).

Table 3. Effect of harsh discipline on behavioral problems

	Behavioral problems							
	Parent report ▫				Child report ‡			
	N	*B*	95% *CI*	R^2	*N*	*B*	95% *CI*	R^2
Harsh discipline (HD) mother								
Model 1 - HD score mother	3,773	**0.34**	0.30, 0.38		3,998	**0.12**	0.08, 0.16	
Model 2$ - HD score mother, adjusted model		**0.24**	0.20, 0.29			**0.10**	0.06, 0.14	
Model 3¥ - HD score mother, fully adjusted model		**0.05**	0.02, 0.09	.36		**0.07**	0.03, 0.11	.05
Harsh discipline (HD) father								
Model 1 - HD score father	3,051	**0.29**	0.24, 0.33		3,172	**0.14**	0.10, 0.18	
Model 2$ - HD score father, adjusted model		**0.21**	0.17, 0.26			**0.11**	0.07, 0.15	
Model 3¥- HD score father, fully adjusted model		**0.08**	0.04, 0.13	.25		**0.09**	0.04, 0.13	.05

Note. Bold typeface represent statistically significant ($p < .05$) associations.
▫ *Z*-score of the Child Behavior Checklist 1,5-5 measured at age 6 ‡ *Z*-score of the Berkeley Puppet Interview measured at age 6. $ Model adjusted for ethnicity of the child, gender of the child, age of the child, number of children in the household, household income, marital status, highest education of the parents, smoking during pregnancy, parental psychopathology and parental report of family functioning ¥ Model additionally adjusted for baseline internalizing/externalizing behavior at age 3 years

To test whether the association of harsh discipline with child self-reported behavioral problems was independent of parent report, we additionally adjusted this relation for *parent* reports of behavioral problems. Both maternal and paternal harsh discipline predicted child reported behavioral problems, independently of parent reported behavioral problems (*B* for maternal harsh discipline = 0.06, *95% CI*: 0.02, 0.11, R^2 = .06; *B* for paternal harsh discipline = 0.07, *95% CI*: 0.02, 0.11, R^2 = .06).

Table 4 shows the relation between harsh discipline and *emotional problems*. Higher levels of maternal and paternal harsh discipline were associated with more parent reported emotional problems (model 3: *B* for maternal harsh discipline = 0.05, *95% CI*: 0.01, 0.08, model 3; *B* for paternal harsh discipline = 0.04, *95% CI*: 0.00, 0.08). Yet, we found that neither maternal nor paternal harsh discipline was related to emotional problems as reported by the child in model 3 (*B* for maternal harsh discipline = 0.03, *95%CI*: -0.02, 0.07; *B* for paternal harsh discipline = 0.02, *95% CI*: -0.02, 0.07).

The above analyses were conducted in the number of children with data available for one or more of the outcome measures to reduce selection bias. Next, we repeated all analyses in those participants with complete data to allow for optimal comparison between analyses. Results were essentially unchanged.

Table 4. Effect of harsh discipline on emotional problems

	Emotional problems							
	Parent report [□]				Child report [ǂ]			
	N	*B*	95% *CI*	R^2	*N*	*B*	95% *CI*	R^2
Harsh discipline (HD) mother								
Model 1 - HD score mother	3,764	**0.24**	0.20, 0.28		4,015	**0.07**	0.03, 0.11	
Model 2[$] - HD score mother, adjusted model		**0.13**	0.09, 0.17			0.03	-0.01, 0.07	
Model 3[¥] - HD score mother, fully adjusted model		**0.05**	0.01, 0.08	.31		0.03	-0.02, 0.07	.06
Harsh discipline (HD) father								
Model 1 - HD score father	3,047	**0.15**	0.10, 0.19		3,182	0.03	-0.01, 0.07	
Model 2[$] - HD score father, adjusted model		**0.08**	0.04, 0.13			0.03	-0.02, 0.07	
Model 3[¥]- HD score father, fully adjusted model		**0.04**	0.00, 0.08	.19		0.02	-0.02, 0.07	.04

Note. Bold typeface represent statistically significant ($p < .05$) associations.
[□] *Z*-score of the Child Behavior Checklist/1,5-5 measured at age 6. [ǂ] *Z*-score of the Berkeley Puppet Interview measured at age 6. [$] Model adjusted for ethnicity of the child, gender of the child, age of the child, number of children in the household, household income, marital status, highest education of the parents, smoking during pregnancy, parental psychopathology and parental report of family functioning. [¥] Model additionally adjusted for baseline internalizing/externalizing behavior at age 3 years

DISCUSSION

Parental harsh discipline -whether used by father or mother- increases the risk of behavioral problems in young children. In the present study, mild forms of harsh parental discipline were negatively associated with parent and child reported behavioral problems. By adjusting for pre-existing problems, we showed that this reflects an increase in problems across a three-year period. Most importantly, we demonstrated that children seem to provide independent information when assessing the effects of parental harsh discipline on behavioral problems, whereas the results for child and parent-reported emotional problems were less consistent.

Studies have repeatedly associated harsh discipline, based on parent reports, with child emotional and behavioral problems (i.e. Larzelere, 2000; Taylor et al., 2010). However, in the present study the effects of harsh discipline on behavioral problems were not restricted to harsh discipline by the father, as proposed by Avakame (1998) and reported by Chang et al. (2003). Rather, maternal harsh discipline had effects very comparable to harsh discipline of the father. Possibly, the disciplining strategies we studied were mild and verbally oriented (e.g. screaming and threatening) and may thus not discriminate well between maternal and paternal discipline. Clear differences between mothers and fathers may be detected only if more extreme forms of harsh discipline are studied. Alternatively, the presence of any harsh behavior in a family is more important than the gender of the disciplining parent. Indeed, partners are often similar in antisocial behavior (Zwirs et al., 2011), that is, mothers who tend to discipline their children harshly more often have partners who also practice this parenting discipline.

Our findings based on child self-reported behavioral problems were not only consistent with those from parent reported behavioral problems, but effects observed using child reports were independent of the parental report. Increased explained variance underpinned this findings. This supports our hypothesis and suggests that children provide unique information on the behavioral consequences of harsh parenting. This observation is clinically relevant since parents using harsh discipline may interpret their children's behavior differently than other parents (Anderson et al., 1986; Chang et al., 2003; Fox et al., 1994; Reid et al., 1987). These biased reports may come about for a number of reasons. For example, parents may report higher levels of child problems in their own defense. Alternatively, highly critical parenting may cause some parents to have a low tolerance for otherwise normal child problems due to stress. Lastly, certain parents tend to notice only the most extreme behaviors (Reid et al., 1987).

Concluding, if all information (on outcome and determinant) is obtained from one informant, reporter bias may occur (Reid et al., 1987). The results from this study indicate that young children may be considered as one of the sources of information in a multi-informant approach on the consequences of harsh parenting. Child self-reports not only confirmed the parental reports, but suggest that scientists may underestimate the effect of harsh parenting as the child provided independent information on the possible behavioral consequences of harsh parenting.

Maternal and paternal harsh discipline were associated with emotional problems as reported by the primary caregiver but not with emotional problems reported by the child. This was in contrast

to rather similar effect sizes observed for the association of harsh parenting with child and parent reported behavioral problems. A relation between harsh discipline and emotional problems in children was hypothesized as children can develop a negative view of the self and feel worthless as a result of a harsh parental discipline style (Sachs-Ericsson, Verona, Joiner, & Preacher, 2006). On the other hand, emotional problems may be determined more by genetic variations and less by role modeling than behavioral problems (Oord, van den, Verhulst, & Boomsma, 1996; Valk, van der, Oord, van den, Verhulst, & Boomsma, 2003). However, some methodological explanations for the discrepancy between results obtained with parent and child reports of emotional problems must be discussed. First, parents distinguished less between emotional and behavioral problems than children. The correlations between parent reported emotional and behavioral problems were higher than those of child reported emotional and behavioral problems (.66 vs .30 in this study). The effect of harsh discipline on parent reported emotional problems may partly reflect the association between harsh discipline and behavioral problems. Second, children may be less accurate reporters of emotional problems (Luby et al., 2007) as these are less concrete aspects of behavior (Harter, 1990). Third, children's ideas about the self do not necessarily match with objectively observable constructs. The concordance between self-perceptions and observable behavior may grow stronger when children age (Davis-Kean et al., 2008). Therefore, we recommend studying the children's perspective on the consequences of harsh discipline over a longer period of time. Finally, these inconsistent results support findings from previous studies suggesting that all informants' reports are imperfect measures of child behavior (Arseneault et al., 2005). Therefore combining information from multiple sources is consdered most optimal (Kraemer et al., 2003).

Strengths and limitations

Some methodological considerations need to be taken into account. Strengths of this study are the large number of population-based participants. In addition, we obtained both parent and child reports of emotional and behavioral problems at age six, and reports of both maternal and paternal harsh discipline were available. Adjusting for baseline problems (pre-existing child problems at age three) allowed us to analyse changes in emotional and behavioral problems in a relatively short period of time.

One of the limitations of our study is that, although we included large numbers of participants, non-response analysis showed some selective attrition. This resulted in an under-representation of children from families with a lower income and mothers without a partner, while families from a low socioeconomic background are at increased risk for both parental harsh discipline (Jansen et al., 2012) and child behavioral problems (Bradley & Corwyn, 2002). However, although prevalence rates have an impact on statistical power, these changes do not necessarily alter the relationship between determinant and outcome (Wolke et al., 2009).

A second limitation is that we had to rely on parent reports only of baseline child problems. Adjusting child self-reported problems for baseline parent reported child problems is not the

optimal adjustment to rule out reverse causality. However, it is not feasible to conduct interviews in three-year old children about their behavior because the BPI and other child self-report instruments only yield reliable estimates in children from older ages (Measelle, Ablow, Cowan, & Cowan, 1998). Although the primary caregiver was asked to fill out the questionnaire assessing child behavior at age six, mostly mothers completed the questionnaire. Therefore, parent reported child emotional and behavioral problems mostly reflected the mothers' views of child problem behavior.

Third, when parents report on their own harsh disciplining, social desirability may lead to a response bias. Even though, in this study only mild forms of harsh discipline were investigated -as three items on the physical assault scale were excluded from the questionnaire (Jansen et al., 2012)- parental underreporting of any verbal or psychological tactics may have been the case. Yet, misclassification in the group of parents that did report harsh disciplining is less likely: if parents reported harsh disciplining tactics, this has most probably been the case. Taken together, these response patterns may have resulted in an underestimation of the effect.

Fourth, emotional and behavioral problems were assessed differently between children and parents as items in the CBCL-questionnaire differ from items in the BPI interview. However, both measures are accepted ways of assessing child emotional and behavioral problems (Ringoot et al., 2013; Rescorla et al, 2007).

Implications

Our study confirmed that even mild forms of harsh parental discipline have substantial effects on the behavioral development of a child. Importantly, this study showed that young children can provide independent, valuable information on behavioral problems as a result of harsh discipline. Although information from young children should be treated with some caution, the possibility to obtain information from very young children provides opportunities for instances when parents are unavailable or unwilling to serve as informants on emotional or behavioral consequences of their parenting behavior. In general, child self-reports could be used in addition to caregiver report when assessing problem behavior, because both the perspective of the child and the parent is important.

The current findings have implications for programs that aim to identify and provide support for children at risk of, or experiencing, harsh discipline. Health care workers should be well aware of the effects of even mild harsh discipline on behavioral problems in children.

REFERENCES

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA Preschool Forms & Profiles.* Burlington: University of Vermont, Research Center for Children, Youth & Families.

Anderson, K. E., Lytton, H., & Romney, D. M. (1986). Mothers' interactions with normal and conduct-disordered boys: Who affects Whom? *Developmental Psychology*, *22*, 604–609.

Arsenault, L., Moffitt, T. E., Caspi, A., Taylor, A., Rijsdijk, F. V., Jaffee, S. R., … Measelle, J. R. (2003). Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. *Journal of Child Psychology and Psychiatry*, *44*, 832–848.

Avakame, E. F. (1998). Intergenerational transmission of violence, self-control, and conjugal violence: a comparative analysis of physical violence and psychological aggression. *Violence and Victims*, *13*(3), 301–316.

Berger, L. M. (2005). Income, family characteristics, and physical violence toward children. *Child Abuse and Neglect*, *29*(2), 107–133.

Bolger, K. E., & Patterson, C. J. (2001). Pathways from child maltreatment to internalizing problems: perceptions of control as mediators and moderators. *Development and Psychopathology*, *13*(4), 913–40.

Boulet, J., & Boss, M. W. (1991). Reliability and validity of the Brief Symptom Inventory. *Psychological Assessment*, *3*(3), 433–437.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–99.

Chang, L., Schwartz, D., Dodge, K. a, & McBride-Chang, C. (2003). Harsh parenting in relation to child emotion regulation and aggression. *Journal of Family Psychology*, *17*(4), 598–606.

Davis-Kean, P. E., Huesmann, L. R., Jager, J., Collins, W. A., Bates, J. E., & Lansford, J. E. (2008). Changes in the relation of self-efficacy beliefs and behaviors across development. *Child Development*, *79*, 1257–1269.

Day, R. D., Peterson, G. W., & Mccracken, C. (1998). Predicting Spanking of Younger and Older Children by Mothers and Fathers, *Journal of Marriage and Family, 60*(1), 79–94.

Fox, R. A., Platz, D. L., & Bentley, K. S. (1994). Maternal factors related to parenting practices, developmental expectations, and perceptions of child behavior problems. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *156*(4), 431–441.

Harter, S. (1990). Adolescent self and identity development. In S. S. Feldman & G. R. Ellio (Eds.), *At the treshold: The developing adolescent* (pp. 352–387). Cambridge, MA: Harvard University Press.

Jaddoe, V. W., van Duijn, C. M., Franco, O. H., van der Heijden, A. J., van IJzendoorn, M. H., de Jongste, J. C., … Hofman, A. (2012). The Generation R Study: design and cohort update 2012. *European Journal of Epidemiology*, *27*(9), 739–56.

Jansen, P. W., Raat, H., Mackenbach, J. P., Hofman, A., Jaddoe, V. W. V., Bakermans-Kranenburg, M. J., … Tiemeier, H. (2012). Early Determinants of Maternal and Paternal Harsh Discipline: The Generation R Study. *Family Relations*, *61*(2), 253–270.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., & Thomas Boyce, W., Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *American Journal of Psychiatry*, *160*, 1566–1577.

Lansford, J. E., Deater-deckard, K., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2004). Ethnic differences in the link between physical discipline and later adolescent externalizing behaviors. *Journal of Child Psychology & Psychiatry*, *45*(4), 801–812.

Lansford, J. E., Dodge, K. A., Malone, P. S., Oburu, P., Palme, K., Bacchini, D., … Quinn, N. (2005). Physical Discipline and Children ' s Adjustment : Cultural Normativeness as a Moderator, *76*(6), 1234–1246.

Larzelere, R. E. (2000). Child outcomes of nonabusive and customary physical punishment by parents: an updated literature review. *Clinical Child and Family Psychology Review*, *3*(4), 199–221.

Luby, J. L., Belden, A., Sullivan, J., & Spitznagel, E. (2007). Preschoolers' contribution to their diagnosis of depression and anxiety: Uses and limitations of young child self-report of symptoms. *Child Psychiatry and Human Development*, *38*(4), 321–338.

McLeod, J. D., Kruttschnitt, C., & Dornfeld, M. (1994). Does parenting explain the effects of structural conditions on children's antisocial behavior? A comparison of Blacks and Whites. *Social Forces*, *73*, 575–604.

Measelle, J. R., Ablow, J. C., Cowan, P. A., & Cowan, C. P. (1998). Assessing Young Children's Views of Their Academic, Social, and Emotional Lives: An Evaluation of the Self-Perception Scales of the Berkeley Puppet Interview. *Child Development*, *69*(6), 1556–1576.

Miller, I.W., Epstein, N.B., Bishop, D.S., Keitner, G. I. (1985). The McMaster Family Assessment Device: reliability and validity. *Journal of Marital and Family Therapy*, *11*(4), 345–356.

Oord, van den, E. J. C. G., Verhulst, F. C., & Boomsma, D. L. (1996). A genetic study of maternal and paternal ratings of problem behaviors in 3-year-old twins. *Journal of Abnormal Psychology*, *105*(3), 349–357.

Prinzie, P., Onghena, P., & Hellinckx, W. (2006). A cohort-sequential multivariate latent growth curve analysis of normative CBCL aggressive and delinquent problem behavior: Associations with harsh discipline and gender. *International Journal of Behavioral Development*, *30*(5), 444–459.

Reid, J. B., Kavanagh, K., & Baldwin, D. V. (1987). Abusive parents' perceptions of child problem behaviors: An example of parental bias. *Journal of Abnormal Child Psychology*, *15*(3), 457–466.

Rescorla, L., Achenbach, T., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., ... & Verhulst, F. (2007). Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *Journal of Emotional and Behavioral Disorders*, *15*(3), 130-142

Ringoot, A.P., Jansen, P. W., Steenweg-de Graaff, J. Measelle, J. R., van der Ende, J., Raat, H., Jaddoe, V. W. V., … Tiemeier, H. (2013). Young children's self-reported emotional, behavioral, and peer problems: The Berkeley Puppet Interview. *Psychological Assessment*, *25*(4), 1273–1285.

Sachs-Ericsson, N., Verona, E., Joiner, T., & Preacher, K. J. (2006). Parental verbal abuse and the mediating role of self-criticism in adult internalizing disorders. *Journal of Affective Disorders*, *93*(1-3), 71–8.

Sidebotham, P., & Golding, J. (2001). Child maltreatment in the "children of the nineties" a longitudinal study of parental risk factors. *Child Abuse and Neglect*, *25*(9), 1177–1200.

Straus, M.A., Hamby, S.L., Moore, D.W., Runyan, D. (1998). Identification of child maltreatment with the Parent-Child Conflict Tactics Scales: development and psychometric data for a national sample of American parents. *Child Abuse and Neglect*, *22*(4), 249–70.

Taylor, C., Manganello, J., Lee, S., & Rice, J. (2010). Mothers' spanking of 3-year-old children and subsequent risk of children's aggressive behavior. *Pediatrics*, *125*(5), e1057–65.

Teicher, M. H., Samson, J. a, Polcari, A., & McGreenery, C. E. (2006). Sticks, stones, and hurtful words: relative effects of various forms of childhood maltreatment. *American Journal of Psychiatry*, *163*(6), 993–1000.

Tick, N. T., Ende, J., van der, Koot, H. M., & Verhulst, F. C. (2007). 14-Year changes in emotional and behavioral problems of Dutch very young children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 1333–1340.

Valk, van der, J. C., Oord, van den, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2003). Genetic and environmental contributions to stability and change in children's internalizing and externalizing problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(10), 1212–1220.

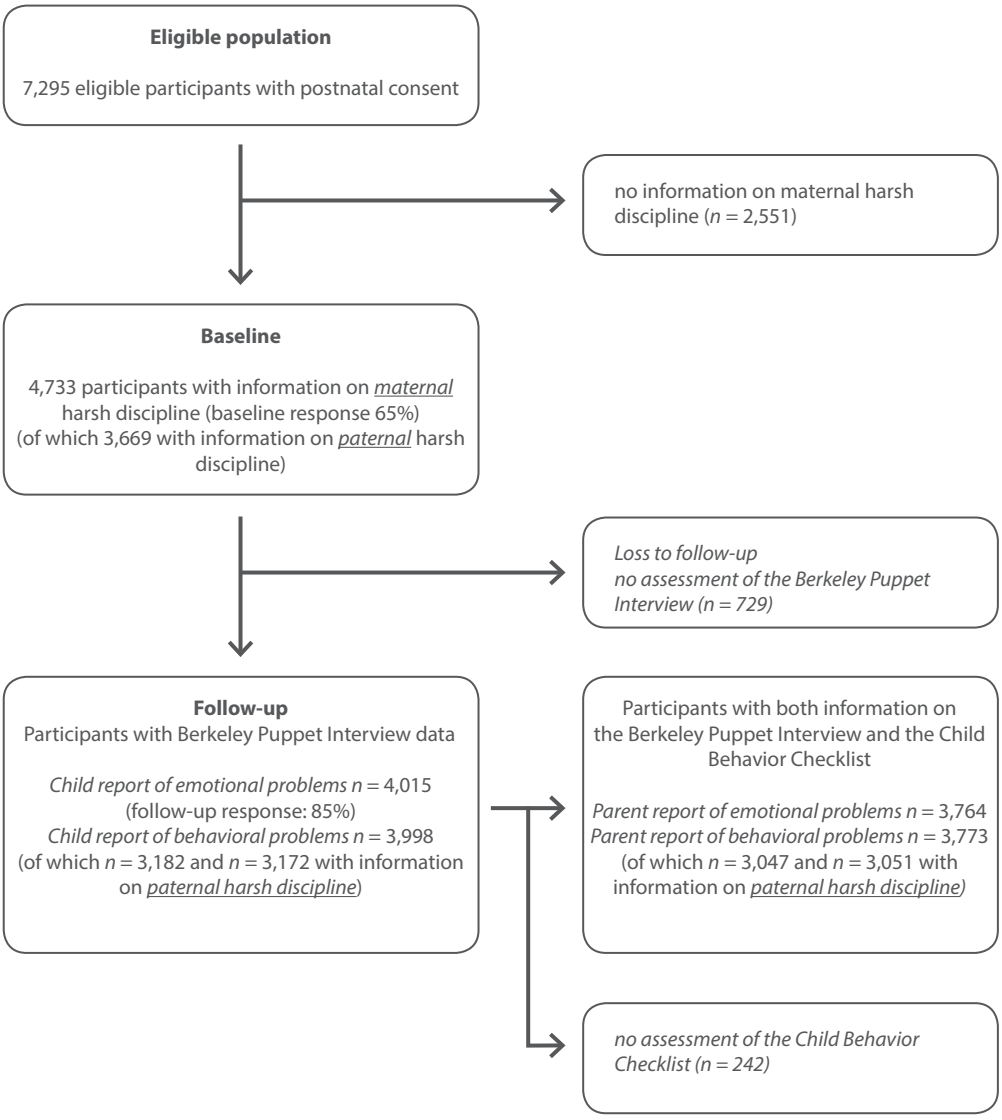
Vostanis, P., Graves, A., Meltzer, H., Goodman, R., Jenkins, R., & Brugha, T. (2006). Relationship between parental psychopathology, parenting strategies and child mental health - findings from the GB national study. *Social Psychiatry and Psychiatric Epidemiology*, *41*(7), 509–14.

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., & Lamberts, K. (2009). Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *British Journal of Psychiatry*, *195*(3), 249–56.

Zwirs, B., Verhulst, F., Jaddoe, V., Hofman, A., Mackenbach, J., & Tiemeier, H. (2011). Partner similarity for self-reported antisocial behaviour among married, cohabiting and dating couples: the Generation R Study. *Psychology, Crime & Law*, *18*(4), 335–349.

SUPPLEMENTARY MATERIAL

Figure S1. Flowchart of study participants
Numbers in the method section give the *n* for emotional problems



chapter 7

Parenting, children's self-regulation and the quality of their peer relationships: The Generation R Study

Ringoot, A.P., Jansen, P.W., Kok, R., Verlinden, M., Jaddoe, V.W.V., Verhulst, F.C., & Tiemeier, H.

To be submitted

chapter 8

General Discussion

DISCUSSION

Accurate diagnosis of young children's problem behavior is of great importance for understanding the early development of psychopathology and for early and effective interventions (Luby, 2010; Webster-Stratton & Taylor, 2001). Children, even young children, may offer a valuable complement to information from the traditional adult informants. Yet, in the Netherlands, no validated instrument was available to systematically assess young children's perspective on their feelings and behavior. To address this gap, the Berkeley Puppet Interview (BPI) was translated into Dutch and introduced in the Generation R Study, a large epidemiological study. The series of studies described in this thesis examined whether information about emotional, behavioral and peer relationship problems, obtained by the Dutch translation of the BPI, provides reliable, valid and valuable information that can be used in research and eventually in clinical settings. The studies also examined the value of young children's reports if implemented along with other outcomes in a multi-informant approach. In this chapter the main findings are summarized and discussed in a broader context. Further, implications for research and practice are presented together with considerations for future studies.

"Out of the mouths of babes": can young children report about problem behavior?

It has long been thought that young children are incapable of reporting about their emotions and behavior. Findings of the current series of studies in a large cohort generally suggested that, if obtained with a structured and developmentally appropriate instrument like the BPI, young children's reports of their own problem behavior can be interpreted validly, provide rather reliable results, are multidimensional and are consistent with, yet independent from, parent reports of child problems.

Construct validity

In the social sciences we are confronted with a difficult situation: we often aim at studying constructs that cannot be assessed directly. As discussed in the introduction, scientists and practitioners therefore rely on information from multiple informants. To obtain information from these informants, questionnaires and interviews are developed with various items that are thought to tap the construct of interest. The construct validity of such measures is an important concept, as it indexes whether results obtained with the questionnaire or interview indeed assess the hypothesized underlying construct (Westen & Rosenthal, 2003). The construct validity of any measure is established by associations with theoretically related constructs (convergent validity), while the measure of interest should not or negatively relate to constructs that are considered to measure something else (divergent validity; Westen & Rosenthal, 2003). Problem behavior has been associated with various factors like socioeconomic disadvantage, parental psychopathology and harsh discipline (Chang et al., 2003; Goodman & Gotlib, 1999; Lavigne et al., 1996). In the

study described in chapter 2, we examined associations between BPI problem scores and known socioeconomic and demographic risk factors that have been described in the literature, supporting the convergent validity of children's self-reports. In addition, the studies described in chapter 5, 6 and 7, provided further support for the validity of the BPI by showing that parental depression, harsh parental discipline, lack of sensitive parenting and self-regulation difficulties were associated with child problems as obtained with the BPI. Our finding that the BPI was associated with known risk factors converges with findings of other studies that indicated that children's reports of behavior, academic competence, and social competence are in line with external criteria like classroom behavior, interpersonal relationships and attachment security (Ablow et al., 1999; Colwell & Lindsey, 2003; Eccles et al., 1993; Goodvin et al., 2008; Harter & Pike, 1984; Ladd, 1990; Marsh et al., 1998; Marsh et al., 2002; Measelle et al., 1998; Sessa et al., 2001).

Additional support for validity can be obtained by examining the relation between one measure with other validated measures. This can be taken as a test of concurrent validity. Yet, in the case of the BPI this is complicated by the fact that other validated measures to assess the same problems in young children rely on a different informant than the child. Consequently, relating children's self-reports to other measures is as much a test of cross-informant relations as a test of concurrent validity. We found that the correlations between children's problem reports and parent report on the Child Behavior Checklist (CBCL) were small (chapter 2). It is known from other studies that self-reports and parent reports generally have modest correlations (Achenbach et al., 1987). Many factors probably contributed to the low agreement we found, such as: the different approaches (interview versus questionnaire); the different scales, items and scoring methods, and; the differing perceptions each informant has about problem behavior (these will be discussed in more detail in the section on 'multi-informant' approaches).

The findings in chapter 3 illustrate that children's responses on the BPI discriminated between children with and without a parent reported Diagnostic Statistical Manual 4th edition (DSM-IV)-based diagnosis, and between children with and without parent reported CBCL problems in the clinical range. Higher BPI externalizing problem scores were also associated with treatment referral two years later, but this association was no longer present after controlling for DSM-based diagnoses obtained from parents. This concurs with other studies indicating that BPI problem levels differed between community and clinically referred children (Ablow et al., 1999), that children with DSM-IV symptoms of conduct disorder scored higher on BPI conduct problems (Arseneault et al., 2005) and that children with a DSM-based diagnosis of anxiety or mood disorders scored higher on core items of depression or anxiety of the BPI (Luby et al., 2007). Other studies also discussed that young children's reports discriminated between a wide range of DSM-based diagnoses (Valla et al., 2002). However, the current series of studies were the first to examine self-reports of children as young as five years.

Dimensionality of children's information

The value of information obtained with the BPI was further examined in chapter 2 by determining whether multiple dimensions could be distinguished in children's report of their problem behavior.

Correlations between various BPI scales supported the idea that young children can distinguish between different types of problem behavior. Each scale had higher correlations with other scales within its designated domain than across domains (e.g. correlations between separation anxiety and overanxiousness were higher than between separation anxiety and conduct problems). Thus, children gave differential information about their emotions and behavior, as they distinguished between different problem types. This is in contrast with a long held believe that young children's self-description are predominantly positive and undifferentiated (Baumeister, 2010). In addition, results from a confirmatory factor analysis (CFA) supported a multidimensional structure of the BPI. This is in line with previous research demonstrating that young children's self-reports may be more differentiated than was long thought (Measelle et al., 1998).

Internal consistency

The internal consistency of a measure, often assessed with Cronbach's alpha, would indicate whether children respond consistently to items within a given scale. While we found that the alphas for the three BPI broadband scales – Internalizing, Externalizing and Peer Relationships – were within an acceptable range, each of the eight separate BPI scales provided only modest alphas. This was particularly true for the depression scale (chapter 2). The heterogeneous items in the scales, in addition to the small number of items per scale, may have affected the estimates (Cortina, 1993). Moreover, there is an ongoing debate whether Cronbach's alpha, the indicator we and many others used, is a useful coefficient (Sijtsma, 2009). It is suggested that in many cases, alpha provides an underestimation of reliability. Further, alpha is often used as an indicator of internal consistency, which makes researchers conclude that if alpha is acceptable the items measure one underlying construct. However, alpha is not necessarily related to internal consistency, as it mainly provides an estimate of the average degree of interrelatedness between the items and as it also depends on the number of items in a scale (Sijtsma, 2009). Therefore, alternatives have been proposed such as the greatest lower bound (GLB) and omega (Revelle & Zinbarg, 2009; Sijtsma, 2009; Stone et al., 2013). In future studies, considering these alternatives may beneficial to obtain better estimates of reliability.

Usefulness of the BPI: a focus on children's perspectives as outcome measure

The existent research on the risk of parental depression and maladaptive parenting for problem behavior in young children primarily used parent reported outcomes (i.e. parental reports of child behavior). The usefulness of the BPI lies for a great part in the opportunity the interview provides to consider child problems from a young child's perspective and to obtain in that way a more nuanced view of child problem behavior.

In chapter 5, the associations of fathers' and mothers' depression with parent and child reported problem behavior in six-year-old children were examined. Previously, children's perspective on well-being was only considered in studies among adolescents. Some of these studies had shown that if children reported problem behavior, fathers' depressive symptoms were more strongly associated with internalizing and externalizing problems than mothers' depressive symptoms (Compas et al.,

1991; Papp, 2012). However, others reported the reverse (Connell & Goodman, 2002; Pearson et al., 2013). The results described in chapter 5 show quite consistently that if young children self-reported their problem behavior, maternal and paternal depression were equally related to internalizing and externalizing problems.

Assessing the young child's perspective also helped better understanding the associations between harsh parenting and child problem behavior. Parents may not always be available to report about their child's behavior or emotions: consider for instance the extreme case of abusive parents. In chapter 6, we did not examine extreme cases, but studied the association between experiences of mild harsh parenting practices and young children's report of problem behavior. Indeed, harsh parental discipline was associated with higher levels of child problem behavior, as has been established by previous studies (Chang et al., 2003; Vostanis et al., 2006). Moreover, young children's reports of behavioral problems provided consistent information with parental information. This suggests that, if reliable adult informants are absent, young children may serve as valuable alternatives.

In chapter 7, harsh and sensitive parenting were studied in relation to children's social behavior. Parenting may influence children's social behavior as young children are thought to transfer the behavioral and family relationship patterns to other social domains like peer relationships (Ladd et al., 2002). Previous studies found that harsh parental discipline is associated with higher levels of parent reported peer aggression in young children, whereas sensitive parenting styles have been associated with lower risks of peer aggression and rejection (Lereya et al., 2013). If child self-reports were used in previous studies, they were mostly obtained from older children. We found that results obtained using young child reported outcomes were consistent with results previously reported in studies using outcomes reported by parents, other adults, or older children. If anything, the results obtained in a subsample using teacher reported peer aggression were attenuated compared to those obtained using information from young children.

Results from the three studies discussed in this paragraph together suggest that young children themselves experience and can report the problems that arise from established parental risk factors.

Usefulness of the BPI: independent information

As discussed above, children's reports obtained by the BPI provided useful information for research, as indexed by consistent associations with caregiver information. Usefulness can also be defined as providing information that is not already reported by other informants (in this case parents) (Arseneault et al., 2005). Even though we could not examine whether children provided unique information over and above information from parents in the prediction of future problem behavior, we could examine whether known correlates of problem behavior were associated with child self-reported problems, independently from parent reported child problems. This issue was addressed in chapters 2 and 6. Associations between demographic and socioeconomic characteristics with BPI problem scores, and associations between harsh parental discipline and BPI externalizing problems, remained after adjustment for parental report of problem behavior. These associations were thus largely independent of parental report of problem behavior and could not be explained with

information obtained from parents only. In contrast, in chapter 3, we found that early elementary school children's reports obtained with the BPI did not predict treatment referral independently from parent reported DSM-based diagnoses.

In summary, if obtained with age appropriate methods like the BPI, information from young children concerning internalizing and externalizing problems can be valid: children's information is associated with known correlates of problem behavior. A disadvantageous family background, harsh parenting and parental depression were all associated with higher levels of child reported problems. Child reported BPI problems were also associated with DSM-based diagnoses and clinical levels of problem behavior as reported by parents, but BPI reports cannot be used to predict future treatment for mental health problems. Further, children's reports appeared to be multi-dimensional, children could discriminate between different types of problems. However, the consistency of items within each scale appeared to be only modest. We also found that associations between known risk factors and children's reports of problem behavior were surprisingly consistent with associations between the same risk factors and parental reports of child problems, and with associations that are known from the literature. Moreover, two of our studies demonstrated that children provided independent information to their parents. Thus, using a child interview, like the BPI, seems to provide researchers with the opportunity to get insight in the way young children perceive their world. Such insights may be particularly valuable in instances when parents are less reliable informants, for example as a result of psychiatric or familial problems. The value of different informants will be further discussed in the next section.

The merits and demerits of a multi-informant approach

Perspectives of various informants often do not converge and this may have important consequences. In a series of studies we compared information from young children with information from caregivers to obtain a better idea of why parents and children disagree when reporting problem behavior, and to examine how information from young children can contribute to our understanding of the development of child problem behavior.

Mother-child disagreement

There is consensus that using information from various sources in the assessment of child problem behavior yields a more accurate picture of the problems (Kraemer et al., 2003). Diverse informants are, among other reasons, needed to tap the cross-situational diversity in children's behavior (Achenbach et al., 1987). Inherent to the use of more than one informant of children's problems is low agreement between the various informants' reports. Contextual factors and informants' perspectives are all thought to contribute to discrepancies (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005). A growing body of literature suggests that informant discrepancies are associated with poor outcomes. Ferdinand et al. (2004) showed that, although agreement that a problem exists signified poorest prognosis, discrepancies between parent and child reports were indicative of poor outcomes. Discrepancies were associated with a larger likelihood of police or judicial contacts,

referral to mental health services, substance abuse and of having externalizing or internalizing problems. Others have also shown that larger parent-child discrepancies concerning a variety of constructs (e.g. parental monitoring, parenting practices, and problem behavior) predicted adverse outcomes like delinquent behavior, internalizing problems and lower social competence, and less parental involvement in treatment (De Los Reyes et al., 2010; Guion et al., 2009; Israel et al., 2007). As a result of informant discrepancies, the choice for a specific reporter can impact prevalence estimates in surveys, the associations found in etiological studies, and even the decisions in clinical practice.

In chapter 4, we studied what factors contribute to discrepancies between children's and mothers' reports of children's problems. Mothers' reports were compared to children's, as mothers are still the most common informant of child problem behavior in research. Mother-child discrepancies were modeled using latent profile analysis in that study. Four profiles were identified, differing in problem level, and in the direction and magnitude of mother-child discrepancies. For instance, one pattern represented a group of children that reported much more problems than their mother (11%), whereas another pattern represented a group of children that reported much less problems than their mothers (13%). Findings suggested that particularly child-related characteristics, like cognitive difficulties, increased the likelihood that children reported more problems than their mothers. In contrast, adverse family and parenting characteristics were the most salient determinants of a group in which mothers reported more problems than their children.

Young children's information as solution to reporter bias and shared-method variance

As discussed, perspectives of different informants often do not converge and this may have important consequences. Depression is one of the factors often thought to contribute to biased reports (Richters, 1992). Information-processing theories suggest that depressed individuals tend to interpret their experiences negatively (Foland-Ross & Gotlib, 2012). Indeed, we found that the strength of associations between paternal and maternal depression with child problems seemed to depend on whether or not parents reported about both depression and child outcomes (chapter 5). If mothers or fathers reported about both their own depression and children's problems, the observed association of the respective parents' depression with child problems seemed inflated. In some instances, estimates were twice as large as compared to estimates obtained by children's reports of their problems. If children reported the outcome, fathers' depression appeared to contribute to young children's problem behavior as much as mothers' depression. Results obtained using one parent's information of both the determinant and outcome thus seemed biased. This may have occurred as a result of distorted information-processing (Foland-Ross & Gotlib, 2012). However, given that parental depression was not measured concurrently with child problems, the inflated estimates probably cannot be entirely explained by this distorted information-processing. Parents with a prior history of depression did not necessarily experience high levels of depressive symptoms while reporting child problems at age six. We concluded that inflated reports resulted from shared-method variance, which may include, but is not restricted to, a 'depressive' bias as discussed above. Shared-method variance is the bias occurring if the same informant is used to

report on determinant and outcome (Collishaw et al., 2009). In the case of shared-method variance, part of the variance between the determinant and the outcome results from characteristics that are constant if the same method or informant is used repeatedly. Our findings demonstrated that future studies of parental depression would benefit from additional information to parents' perspectives to reduce biases, with possibly children as one of these information sources.

Findings in chapter 4 and 5 seem to contradict each other in one aspect. In chapter 4, parental depression appeared to be no independent contributor to mother-child discrepancies. Other determinants than parental depression were more strongly associated to reported discrepancies. Yet, results in chapter 5 suggest that, if examining the effect of depression, it is important to consider that estimates may be inflated once the person reporting depression is the same as the person reporting (child) outcomes. This difference in findings most likely resulted from the different variables that were controlled for in both studies. In the study on mother-child discrepancies we controlled the analyses for family functioning, however, as family functioning is strongly associated with parental psychopathology (e.g. Dickstein et al., 1998; Low & Stocker, 2005), this may have attenuated the effect of depression. To some extent controlling for interdependent factors may even be considered overadjustment, as family functioning and parental psychopathology are closely intertwined and may determine each other.

Another factor that can contribute to biased parent reports is harsh parenting. Parents who tend to use harsh discipline practices, may perceive their children differently (Reid et al., 1987), and may consequently report more problem behavior. In our studies we also examined whether parent and child reports yielded similar results in a study of harsh parental discipline and child problem behavior (chapter 6). We found that parental harsh discipline -whether used by father or mother- increased the risk of behavioral problems in young children. Like in the study examining parental depression (chapter 5), the (unadjusted) associations were stronger and may have been inflated if parents reported both harsh parenting and child outcomes. After adjustment for parent reported pre-existing child problems at age three, estimates for the effect of harsh parenting on children's problems were similar across parent and child reported outcomes. Most importantly, the results indicated that child reports not only confirmed the findings obtained using parent reports (i.e. effect estimates of fully adjusted models were similar across parent and child reported outcomes), but results also indicated that children provided independent information. If relying on parent reports only, researchers may underestimate the effect of harsh parenting.

In general, results discussed in this paragraph highlight that young children's self-reports can help to address potential bias in parent reports.

A closer look at other 'forgotten' informants

In the studies of associations between important parental factors and child well-being described in chapter 5, 6 and 7, we also focused on fathers who, like young children, only recently received more attention of researchers (Ramchandani & Psychogiou, 2009). Fathers' depression and fathers' harsh parenting were included as determinant in addition to depression and harsh parenting of mothers. Overall, father depression and father harsh parenting seemed to influence child outcomes just like

the maternal factors. Yet, the mechanism through which father's harsh discpline is associated with children's peer relations, may differ from that of mothers (chapter 7). Even if the assocations between maternal and paternal risk factors and child outcomes seem similar, underlying mechanisms can be different. In general fathers are thought to play a different role in children's upbringing (Lewis & Lamb, 2003), with fathers encouraging autonomy and self-control through play (i.e. rough-and-tumble play), which may teach children to cope with aggressive impulses and anxiety for new situations (Paquette, 2004). Although mechanism may be different, a general conclusion of our studies was that fathers', like mothers', depression and harsh parenting seems to play an important role in the healthy development of their children.

In summary, multi-informant reports almost inevitably result in discrepancies. These discrepancies can partly be explained by environmental factors that may impact the perspectives of informants, like adverse family functioning and harsh parental discipline. If the same informant provides information on environmental determinants and the outcome, this may result in biased effect estimates. Although a demerit of a multi-informant approach is that convergence between information from different reporters is hardly ever conceived, the merit of comparing multi-informant data is that each informant can also provide valuable information, which can help to limit potential biases.

Methodological considerations

Data aggregation if using multi-informant approaches

Studies throughout the past decades, including those presented in the current thesis, have demonstrated that there is no such thing as the optimal informant of child problem behavior. The solution therefore is to use multiple informants, of which young children may be one. However, the use of multiple informants also confronts researchers with the question how, if desired, to aggregate multi-informant data.

Primarily it is important to consider that there are two ways to look at multi-informant data. De Los Reyes et al. (2013) discuss that on the one hand there are researchers assuming that by aggregating data from multiple informants, a better estimator of the true problem is obtained. In their view discrepancies are considered a nuisance (Roberts & Caspi, 2001). On the other hand, scholars state that because of differing perspectives the information from various informants cannot be used interchangeably, but that each informant may provide valid information (Hunsley & Mash, 2011). Most likely, both approaches can exist alongside each other. In this section I will discuss practical consideration to decide what approach to use when. A more theoretical discussion of the different ways to perceive multi-informant data is provided by De Los Reyes et al. (2013).

There are three ways often applied to analyze multi-informant data (De Los Reyes et al., 2013; Goldwasser & Fitzmaurice, 2001). A first approach is to analyze data from each informant separately. Often, information from one informant is considered the primary outcome and a second informant the secondary outcome. This approach was selected in the studies described in chapter 6 and 7. One reason to choose a primary outcome is that for the secondary outcome less extensive data

are available or that the data are only available in a subsample, the latter was the case in chapter 7. In chapter 6 we chose to perform analyses in each informant separately, because this was one of the first studies examining the association between harsh parental discipline and children's self-report of problem behavior. Therefore, we were specifically interested in the children's reports. A disadvantage of this approach is, however, that if results across informants differ, they cannot easily be compared. Also, if the different outcome analyses were based on slightly different subsets of the data, this further complicates comparison. In addition, studying results for each informant separately increases the risk of false positive results (type I errors), as this risk is present for each of the associations studied.

A second approach to multi-informant data is to pool data by using 'and' or 'or' rules. Using the 'and' rule implies that both of two informants indicate that a problem is above a threshold level. In contrast, if the 'or' rule is followed, only one of two informants needs to indicate that a problem is above a certain threshold. Such an approach can be applied if data are dichotomous. In the case of continuous data, pooling can be performed by taking the standardized mean of each informant's report. The advantage of this pooling approach is that it is straightforward. However, an important disadvantage is that we can no longer discover if the association between a determinant and outcome differed between informants. In chapter 5, we discussed that the only available studies testing the association between parental depression and young children's problems using multi-informant data of the outcome had used mean pooling or had generated a diagnosis applying the 'or' rule (Gere et al., 2013; Ramchandani et al., 2008a; Ramchandani et al., 2008b). Consequently, the contribution of each specific informant could not be disentangled. Potentially, one source was driving the association, but this could not be ruled out.

A third approach that is often used is to combine multiple measures of a single behavior into latent constructs, thus to obtain a consensus measure. This can be done in various ways and has been applied to studies including BPI data (e.g. structural equation modeling (SEM), factor analysis or principal component analysis). Often, these techniques focus on examining the shared variance among multiple informants' reports of the same behavior (Arseneault et al., 2003; De Los Reyes et al., 2013; Gere et al., 2013; Perren et al., 2006). Here I will highlight one method to obtain a consensus measure. Kraemer et al. (2003) pointed out in a paper on methods to combine information from multiple sources that if information from three sources (like children, parents and teachers) are triangulated, using principle component analyses, three factors can be obtained. A first factor thought to represent the 'trait' (i.e. a relatively bias free estimate of the problem), a second context factor (contrasting home vs. non-home), and a third perspective factor (contrasting adults vs. child). This method has been used to obtain a multi-informant score that can be applied in studies both as a determinant (Perren et al., 2006) and as an outcome (Obradović et al., 2010).

Aggregating multi-informant data may not be appropriate in all circumstances. It is important to carefully consider the data, before determining whether or not to aggregate them. Before pooling or aggregating data, it may be appropriate to demonstrate with analytical models that various informants in fact reported about the same problem. This can be tested using multivariate linear regression or generalized-estimating-equations (GEE) (Goldwasser & Fitzmaurice, 2001). In

a multivariate linear regression model, the information obtained from multiple informants can be treated as if they are repeated measures; they can be modeled simultaneously, thereby taking into consideration the correlation between informant reports. An interaction term between the determinant and informant can be included in the model to test whether associations between determinant and outcome depend on the informant (Goldwasser & Fitzmaurice, 2001). If a significant interaction is absent, the interaction term may be dropped and an overall estimate can be interpreted. This method was applied in the study described in chapter 5. In that chapter, we had to conclude that analyzing aggregated data would be inappropriate. There was a significant interaction between parents' depression and the informant of child problems, indicating that the size and direction of associations between parental depression and child problems depended on the informant of child problems. Therefore, in the next part of that study, associations were examined separately.

A circumstance in which aggregating data can be appropriate, is if chances are low that information from two informants yields diverging outcomes - for example if the studied determinant is not likely affected by social desirable responding (e.g. neuroanatomy) or if shared-method variance is unlikely- and if the various informant measures are known to be sufficiently valid and reliable. Data aggregation can be performed using a multivariate regression model discussed above. Alternatively, if a more simple method is preferred, pooling data can also be considered by using the average of two or more z-standardized informant scores. In one of our studies (not included in this thesis) on the association between neuroanatomical correlates and child and parent reported peer aggression, we applied this straightforward method of data aggregation (Thijssen et al., 2015). Interestingly, the obtained aggression score correlated almost perfectly (r = .93) with the 'trait' factor from the more advanced aggregation strategy discussed by Kraemer et al. (2003). This has also been discussed by Noordhof et al. (2008): all factors from the approach discussed by Kraemer et al. (2003) can be derived by manual computations.

In summary, if one uses multi-informant reports and decides whether or not to aggregate the data from multiple sources, the first step is to test whether informants report on the same underlying construct by multivariate regression models. This is particularly advisable if a construct is investigated that may be prone to reporter bias. In case associations depend on the informant, analyzing reports from each informant separately and considering factors that can explain the discrepancies is the most appropriate strategy. Yet, if associations are independent of the informant, data can be aggregated by methods such as multivariate regression, by including multiple informants in SEM models, or by using a consensus measure.

Factors affecting children's reports

In the introduction of this thesis, several factors were discussed that were taken into account in the development of the BPI. Nevertheless, several potential biases in children's reports remain to be considered. These relate to cognitive and contextual influences on interview responses.

Cognitive processes influence the way people respond to interview questions (Brener et al., 2003; Cannell et al., 1981). Basic models describing response process indicate that this is influenced by:

1) comprehension of the question; 2) retrieving the requested information from memory, which depends on prior experiences and beliefs, 3) decision-making (i.e. does the response fulfill the objective of the question) and 4) response generation (Brener et al., 2003).

In the development of the BPI, attempts have been made to match the sentence structure and language to the developmental level of young children (Measelle et al., 1998). Consequently, errors in the first process, 'comprehension', and in the third or fourth processes may not be very likely to occur. Nevertheless, if requested information becomes more specific or complex, the more likely it is that reports are less accurate in young children (Eder, 1989). This may have affected children's reports, particularly their reports of internalizing problems. Throughout the studies discussed in the present thesis, children seemed to have more difficulties to report about internalizing problems: the alpha of the BPI depression scale was lower than that of other scales (chapter 2); parent reported DSM-based diagnoses were less strongly related to children's reports of internalizing problems than externalizing problems (chapter 3); and harsh parenting was associated with children's internalizing problems if parents reported, but not if children self-reported internalizing problems (chapter 6). These observations can be explained by the notion that reporting about internalizing problems might be difficult for children, as this requires introspective abilities to a larger extent than reporting about externalizing problems. Young children may have had problems comprehending items concerning internalizing problems, and may be better at reporting concrete and core aspects of behavior (Harter, 1990; Luby et al., 2007). Alternatively, not only young children themselves find it difficult to report about their internalizing problems. Various studies demonstrated that parents and particularly teachers also have difficulties to report about children's internalizing problems (Achenbach et al., 1987; Wu et al., 1999).

The second process, information processing, may have been most vulnerable to errors. During the BPI, children were asked to respond to statements like 'I am a happy kid' versus 'I am not a happy kid'. It is conceivable that responses to questions like these depended on children's mood or on what happened right before they entered the examination room. If for example right before the interview children heard that their best friend would move, their pet was ill, or if they lost a toy, these circumstances could have impacted children's response tendencies. In addition, even if children were not influenced by experiences immediately prior to the interview, the absence of a fixed reporting period in the BPI may have impacted correlations of BPI scales with the CBCL or DISC-YC, because these methods use fixed reporting periods. Nevertheless, modest but significant association were found in chapter 2 and 3 between children's reports of internalizing and externalizing problems and parent reported DSM-based diagnoses, as well as CBCL problem scores.

Other factors that potentially influenced the validity of children's responses are issues associated with the situation in which children were interviewed. The presence of the interviewer and also of a caregiver in the room (mostly a parent), even though he or she was occupied in a different task and behind a curtain, may have influenced the child's perception of privacy and consequently may have resulted in social desirable responses (Cannell et al., 1981). Social desirability may have played a role in chapter 7, where children reported on peer aggression behavior. Children know that teasing and hitting other kids is generally not accepted. They may thus have underreported

these behaviors. However, even though social desirability could have played a role, associations were found between parenting and children's peer aggression. If social desirability played a role in children's responses, this most likely resulted in an underestimation of associations.

Other potential limitations

Several other issues must also be considered when interpreting the findings of the studies in his thesis. First, child reported outcomes were measured only once. Consequently, reversed causality (i.e. child problem behavior influencing parenting or parental depression instead of the other way around) cannot be ruled out in some of our studies. However, it is not feasible to conduct interviews in children under four years of age about their behavior, because the BPI and other child self-report instruments only yield reliable estimates in children from older ages (Measelle et al., 1998). Second, although we examined a range of validity and reliability indicators, we did not obtain test-retest reliability estimates. The children visited the research center only once, which precluded obtaining test-retest reliability estimates. Other studies using the BPI in community samples, have demonstrated moderate to high test–retest reliability if collected several days apart. Test-retest correlations ranged from .43 to .86 on the scales used in the present study (Ablow & Measelle, 2003; Ablow et al., 1999). In another study, the test-retest reliability with a one year time interval ranged from .26 to .39 (Stone et al., 2014). Differing test-retest correlations across the two studies likely reflect changes in behavior and affect over time. Third, given that the data collection in the Generation R Study is still ongoing, we did not obtain information on the BPI's predictive validity, while this is an important indicator of a method's validity. More information on predictive validity could be obtained in future studies.

Implications for research and practice

In the current series of studies we introduced the BPI as a measure to obtain standardized and age-appropriate information from young children about problem behavior. The findings of the current series of studies provide primarily implications for research. Findings implicate that in a multi-informant approach, information from young children can be added. Particularly if studying associations between determinants that are prone to reporter biases like parental depression or harsh discipline, or if a researcher anticipates that shared-method variance could influence study results, adding another informant like the young child can solve such problems. Nevertheless, it should be kept in mind that associations between BPI problem scores and other correlates of problem behavior were generally small. In addition, administering the interview and coding the interview is laborious. Still, the BPI seems promising for research in young children. Given the limited or biased perspectives of each informant, given the potential of spurious associations if both determinant and child problem behavior are assessed by the same adult informant, and given that the response-rate from additional informants like teachers is often low and thus limiting the power of a study, adding an additional source of information may become a necessary good. The results of the current series of studies indicate that young children can be considered as one of these information sources in research.

The findings suggest that the BPI is a relatively sound and feasible instrument to apply in studies with young elementary school children. This may also have implications for screening in youth care. By disregarding young children's perspectives, some problems can remain unnoticed. Consequently, one intervenes only once problems become worse or constitute a burden for the entire family. Timely identification of psychopathology may be of importance to prevent persistence of problem behavior into older age (Essex et al., 2009; Hofstra et al., 2000, 2002). Interviewing elementary school children with the BPI could be a way to access their inner world in a playful manner and to get an idea of potential problems requiring further investigation. If considering applying the BPI for screening purposes, it must be kept in mind that the BPI did not predict treatment referral or DSM-based diagnoses in these young children. Rather, the BPI should be considered as a method to get a first hint of potential problems and should be used in addition to other information.

To really extend the use of the BPI beyond research applications, future studies should examine the experiences of clinicians when applying the BPI in a clinical setting. Little is known about the use of the BPI for clinical purposes. As yet, results are promising because we found that children's reports are associated with DSM-based diagnoses, and there are no indications that the value of children's reports would be different in a clinical sample than in the current population based sample. Instead, the BPI may even perform better in a clinical sample (Ablow et al., 1999). However, more thorough investigation of the BPI is needed before it can be implemented for clinical use. To have meaningful value for clinicians, there should be evidence that information obtained with the BPI adds to knowledge from already available diagnostic procedures (Hunsley & Mash, 2011). Although findings suggested that the BPI provides information independent of mother report, longitudinal studies are needed. This is particularly important in the current era in which evidence based assessment and treatment are the norm (Hunsley & Mash, 2011). Future studies may aim at examining whether the use of a child interview in a diagnostic process gives unique information about child-functioning that cannot be provided by other measures, and importantly whether the use of a child interview predicts favorable treatment outcomes. Thus, although the BPI seems a promising instrument, it cannot be applied in clinical settings yet. More research is needed on the incremental value of the BPI. However, given the relatively small associations found in most of the studies described in this thesis -even though associations are generally stronger in clinical samples than in the general population samples- I do not expect that the BPI will provide clinically relevant incremental validity. Most likely it should not be added to standard diagnostic batteries. Nevertheless, the BPI may be a very valuable instrument for individual cases, in which parents are unable or unwilling to provide information, if referral seems to be predominantly based on parents' instead of children's disease burden or the other way around, if parents may under-report their children's problems to avoid stigmatization, or if it seems difficult to obtain insight in children's inner worlds. In these instances obtaining standardized information from young children may prove valuable to practitioners.

To summarize, the BPI alone cannot be used as a screening instrument for psychiatric diagnoses and future research on the clinical implications of the BPI is recommended. However, results of the current series of studies also suggest that the BPI is a sound measure to obtain self-reports from

children between four and eight years, provided that it is used in concert with other measures, or that it is used to test specific hypotheses about child problem behavior. If applied as such, the BPI may enable researchers to gain more accurate insights in child problem behavior, and it may enhance their ability to understand factors relevant to the early development of problem behavior.

REFERENCES

Ablow, J.C., & Measelle, J.R. (2003). *Manual for the Berkeley Puppet Interview: Symptomatology, Social, and Academic Modules (BPI 1.0)*. MacArthur foundation Research Network Psychopathology and Development (David J. Kupfer, Chair): The MacArthur Working Group on Outcome Assessment

Ablow, J.C., Measelle, J.R., Kraemer, H.C., Harrington, R., Luby, J., Smider, N., et al. (1999). The MacArthur Three-City Outcome Study: Evaluating Multi-Informant Measures of Young Children's Symptomatology. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(12), 1580-1590.

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213-232.

Arseneault, L., Kim-Cohen, J., Taylor, A., Caspi, A., & Moffitt, T. (2005). Psychometric Evaluation of 5- and 7-Year-Old Children's Self-Reports of Conduct Problems. *Journal of Abnormal Child Psychology, 33*(5), 537-550.

Arseneault, L., Moffitt, T.E., Caspi, A., Taylor, A., Rijsdijk, F.V., Jaffee, S.R., et al. (2003). Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. *Journal of Child Psychology and Psychiatry, 44*(6), 832-848.

Baumeister, R.F. (2010). The self. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of science* (pp. 139-175): Oxford University Press.

Brener, N.D., Billy, J.O.G., & Grady, W.R. (2003). Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: evidence from the scientific literature. *Journal of Adolescent Health, 33*(6), 436-457.

Cannell, C.F., Miller, P.V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*(4), 389-437.

Chang, L., Schwartz, D., Dodge, K.A., & McBride-Chang, C. (2003). Harsh parenting in relation to child emotion regulation and aggression. *Journal of Family Psychology, 17*(4), 598.

Collishaw, S., Goodman, R., Ford, T., Rabe-Hesketh, S., & Pickles, A. (2009). How far are associations between child, family and community factors and child psychopathology informant-specific and informant-general? *Journal of Child Psychology and Psychiatry, 50*(5), 571-580.

Colwell, J.M., & Lindsey, E.W. (2003). Teacher-Child Interactions and Preschool Children's Perceptions of Self and Peers. *Early Child Development and Care, 173*(2-3), 249-258.

Compas, B., Phares, V., Banez, G., & Howell, D. (1991). Correlates of internalizing and externalizing behavior problems: Perceived competence causal attributions, and parental symptoms. *Journal of Abnormal Child Psychology, 19*(2), 197-218.

Connell, A.M., & Goodman, S.H. (2002). The association between psychopathology in fathers versus mothers and children's internalizing and externalizing behavior problems: a meta-analysis. *Psychological bulletin, 128*(5), 746.

Cortina, J.M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

De Los Reyes, A., Goodman, K.L., Kliewer, W., & Reid-Quinones, K. (2010). The longitudinal consistency of mother-child reporting discrepancies of parental monitoring and their ability to predict child delinquent behaviors two years later. *Journal of Youth and Adolescence, 39*(12), 1417-1430.

De Los Reyes, A., & Kazdin, A.E. (2005). Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483-509.

De Los Reyes, A., Thomas, S.A., Goodman, K.L., & Kundey, S.M.A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9*, 123-149.

Dickstein, S., Seifer, R., Hayden, L.C., Schiller, M., Sameroff, A.J., Keitner, G., et al. (1998). Levels of Family Assessment: II. Impact of Maternal Psychopathology on Family Functioning. *Psychological Assessment, 12*(1), 23-40.

Eccles, J., Wigfield, A., Harold, R.D., & Blumenfeld, P. (1993). Age and Gender Differences in Children's Self- and Task Perceptions during Elementary School. *Child Development, 64*(3), 830-847.

Eder, R.A. (1989). The emergent personologist: The structure and content of 3½-, 5½-, and 7½-year-olds' concepts of themselves and other persons. *Child Development*, 1218-1228.

Essex, M.J., Kraemer, H.C., Slattery, M.J., Burk, L.R., Boyce, W.T., Woodward, H.R., et al. (2009). Screening for childhood mental health problems: outcomes and early identification. *Journal of Child Psychology & Psychiatry, 50*(5), 562-570.

Ferdinand, R.F., van der Ende, J., & Verhulst, F.C. (2004). Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology, 113*(2), 198.

Foland-Ross, L.C., & Gotlib, I.H. (2012). Cognitive and neural aspects of information processing in major depressive disorder: an integrative perspective. *Frontiers in Psychology, 3*.

Gere, M.K., Hagen, K.A., Villabø, M.A., Arnberg, K., Neumer, S.-P., & Torgersen, S. (2013). Fathers' mental health as a protective factor in the relationship between maternal and child depressive symptoms. *Depression and Anxiety, 30*(1), 31-38.

Goldwasser, M.A., & Fitzmaurice, G.M. (2001). Multivariate linear regression analysis of childhood psychopathology using multiple informant data. *International Journal of Methods in Psychiatric Research, 10*(1), 1-10.

Goodman, S.H., & Gotlib, I.H. (1999). Risk for psychopathology in the children of depressed mothers: a developmental model for understanding mechanisms of transmission. *Psychological Review, 106*(3), 458.

Goodvin, R., Meyer, S., Thompson, R.A., & Hayes, R. (2008). Self-understanding in early childhood: associations with child attachment security and maternal negative affect. *Attachment and Human Development, 10*(4), 433-450.

Guion, K., Mrug, S., & Windle, M. (2009). Predictive value of informant discrepancies in reports of parenting: Relations to early adolescents' adjustment. *Journal of Abnormal Child Psychology, 37*(1), 17-30.

Harter, S. (1990). Developmental Differences in the Nature of Self-Representations: Implications for the Understanding, Assessment, and Treatment of Maladaptive Behavior. *Cognitive Therapy and Research, 14*(2), 113-142.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55*(6), 1969-1982.

Hofstra, M.B., Van der Ende, J., & Verhulst, F.C. (2000). Continuity and change of psychopathology from childhood into adulthood: a 14-year follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry., 39*(7), 850-858.

Hofstra, M.B., van der Ende, J., & Verhulst, F.C. (2002). Child and adolescent problems predict DSM-IV disorders in adulthood: a 14-year follow-up of a Dutch epidemiological sample. *Journal of the American Academy of Child & Adolescent Psychiatry., 41*(2), 182-189.

Hunsley, J., & Mash, E.J. (2011). Evidence-based assessment. *The Oxford handbook of clinical psychology*, 76-97.

Israel, P., Thomsen, P.H., Langeveld, J.H., & Stormark, K.M. (2007). Parent–youth discrepancy in the assessment and treatment of youth in usual clinical care setting: consequences to parent involvement. *European Child & Adolescent Psychiatry, 16*(2), 138-148.

Kraemer, H.C., Measelle, J.R., Ablow, J.C., Essex, M.J., Boyce, W.T., & Kupfer, D.J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*(9), 1566-1577.

Ladd, G.W. (1990). Having Friends, Keeping Friends, Making Friends, and Being Liked by Peers in the Classroom: Predictors of Children's Early School Adjustment? *Child Development, 61*(4), 1081-1100.

Ladd, G.W., Pettit, G.S., & Bornstein, M.H. (2002). Parenting and the development of children's peer relationships. *Handbook of Parenting. Vol. 5: Practical Issues in Parenting*, 269-309.

Lavigne, J.V., Gibbons, R.D., Christoffel, K.K., Arend, R., Rosenbaum, D., Binns, H., et al. (1996). Prevalence rates and correlates of psychiatric disorders among preschool children. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(2), 204-214.

Lereya, S.T., Samara, M., & Wolke, D. (2013). Parenting behavior and the risk of becoming a victim and a bully/victim: A meta-analysis study. *Child Abuse & Neglect, 37*(12), 1091-1108.

Lewis, C., & Lamb, M.E. (2003). Fathers' influences on children's development: The evidence from two-parent families. *European Journal of Psychology of Education, 18*(2), 211-228.

Low, S.M., & Stocker, C. (2005). Family Functioning and Children's Adjustment: Associations among parents' depressed mood, marital hostility, parent-child hostility, and children's adjustment. *Journal of Family Psychology, 19*(3), 394-403.

Luby, J.L. (2010). Preschool Depression The importance of identification of depression early in development. *Current directions in psychological science, 19*(2), 91-95.

Luby, J.L., Belden, A., Sullivan, J., & Spitznagel, E. (2007). Preschoolers' contribution to their diagnosis of depression and anxiety: uses and limitations of young child self-report of symptoms. *Child Psychiatry and Human Development, 38*(4), 321-338.

Marsh, H.W., Craven, R., & Debus, R. (1998). Structure, stability, and development of young children's self-concepts: a multicohort-multioccasion study. *Child Development, 69*(4), 1030-1053.

Marsh, H.W., Ellis, L.A., & Craven, R.G. (2002). How do preschool children feel about themselves? Unraveling measurement and multidimensional self-concept structure. *Developmental Psychology, 38*(3), 376-393.

Measelle, J.R., Ablow, J.C., Cowan, P.A., & Cowan, C.P. (1998). Assessing young children's views of their academic, social, and emotional lives: an evaluation of the self-perception scales of the Berkeley Puppet Interview. *Child Development, 69*(6), 1556-1576.

Noordhof, A., Oldehinkel, A.J., Verhulst, F.C., & Ormel, J. (2008). Optimal use of multi-informant data on co-occurrence of internalizing and externalizing problems: the TRAILS study. *International Journal of Methods in Psychiatric Research, 17*(3), 174-183.

Obradović, J., Bush, N.R., Stamperdahl, J., Adler, N.E., & Boyce, W.T. (2010). Biological sensitivity to context: The interactive effects of stress reactivity and family adversity on socioemotional behavior and school readiness. *Child Development, 81*(1), 270-289.

Papp, L. (2012). Longitudinal Associations Between Parental and Children's Depressive Symptoms in the Context of Interparental Relationship Functioning. *Journal of Child and Family Studies, 21*(2), 199-207.

Paquette, D. (2004). Theorizing the Father-Child Relationship: Mechanisms and Developmental Outcomes. *Human Development, 47*(4), 193-219.

Pearson, R.M., Evans, J., Kounali, D., & et al. (2013). Maternal depression during pregnancy and the postnatal period: Risks and possible mechanisms for offspring depression at age 18 years. *JAMA Psychiatry, 70*(12), 1312-1319.

Perren, S., von Wyl, A., Stadelmann, S., Bürgin, D., & Von Klitzing, K. (2006). Associations between behavioral/emotional difficulties in kindergarten children and the quality of their peer relationships. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*(7), 867-876.

Ramchandani, P., O'Connor, T., Evans, J., Heron, J., Murray, L., & Stein, A. (2008a). The effects of pre- and postnatal depression in fathers: a natural experiment comparing the effects of exposure to depression on offspring. *Journal of Child Psychology and Psychiatry, 49*(10), 1069-1078.

Ramchandani, P., & Psychogiou, L. (2009). Paternal psychiatric disorders and children's psychosocial development. *The Lancet, 374*(9690), 646-653.

Ramchandani, P., Stein, A., O'Connor, T.G., Heron, J.O.N., Murray, L., & Evans, J. (2008b). Depression in Men in the Postnatal Period and Later Child Psychopathology: A Population Cohort Study. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(4), 390-398.

Reid, J.B., Kavanagh, K., & Baldwin, D.V. (1987). Abusive parents' perceptions of child problem behaviors: An example of parental bias. *Journal of Abnormal Child Psychology, 15*(3), 457-466.

Revelle, W., & Zinbarg, R.E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika, 74*(1), 145-154.

Richters, J.E. (1992). Depressed mothers as informants about their children: a critical review of the evidence for distortion. *Psychological Bulletin, 112*(3), 485-499.

Roberts, B.W., & Caspi, A. (2001). Personality Development and the Person-Situation Debate: It's Déjà Vu All Over Again. *Psychological Inquiry, 12*(2), 104-109.

Sessa, F.M., Avenevoli, S., Steinberg, L., & Morris, A.S. (2001). Correspondence among informants on parenting: preschool children, mothers, and observers. *Journal of Family Psychology, 15*(1), 53-68.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107-120.

Stone, L.L., Otten, R., Ringlever, L., Hiemstra, M., Engels, R.C.M.E., Vermulst, A.A., et al. (2013). The parent version of the strengths and difficulties questionnaire. *European Journal of Psychological Assessment, 29*(1), 44-50.

Stone, L.L., van Daal, C., van der Maten, M., Engels, R.C.M.E., Janssens, J.M.A.M., & Otten, R. (2014). *The Berkeley Puppet Interview: A Screening Instrument for Measuring Psychopathology in Young Children.* Paper presented at the Child & Youth Care Forum.

Thijssen, S., Ringoot, A.P., Wildeboer, A., Bakermans-Kranenburg, M.J., el Marroun, H., Hofman, A., Jaddoe, V.W.V., Verhulst, F., Tiemeier, H., van IJzendoorn, M.H., & White, T. (2015). Brain morphology of childhood aggressive

behavior: A multi-informant study in school-age children. *Cognitive, Affective, & Behavioral neuroscience*. Advance online publication.

Valla, J.P., Kovess, V., Chee, C.C., Berthiaume, C., Vantalon, V., Piquet, C., et al. (2002). A French study of the Dominic Interactive. *Social Psychiatry and Psychiatric Epidemiology, 37*(9), 441-448.

Vostanis, P., Graves, A., Meltzer, H., Goodman, R., Jenkins, R., & Brugha, T. (2006). Relationship between parental psychopathology, parenting strategies and child mental health. *Social Psychiatry and Psychiatric Epidemiology, 41*(7), 509-514.

Webster-Stratton, C., & Taylor, T. (2001). Nipping early risk factors in the bud: Preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8 years). *Prevention Science, 2*(3), 165-192.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personalitty and Social Psychology, 84*(3), 608.

Wu, P., Hoven, C.W., Bird, H.R., Moore, R.E., Cohen, P., Alegria, M., et al. (1999). Depressive and disruptive disorders and mental health service utilization in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 38*(9), 1081-1090.

SUMMARY

samenvatting

SUMMARY

Researchers in the field of psychology and psychiatry are faced with a difficult situation when assessing problem behavior in young children. Often, the aim is to measure constructs that cannot be measured directly, nor with an objective tool. Consequently, interviews and questionnaires are developed that are preferably administered to multiple informants to obtain the most accurate description of the studied problem. A particularly difficult situation pertains to the assessment of psychological problems in young children. Rather than asking children themselves about their feelings and behavior, it has long been common practice to obtain information solely from parents (mainly mothers), teachers and observations. There are only a few validated age-appropriate methods to systematically obtain information on a wide range of emotional and behavioral problems as reported by children themselves, and in the Netherlands there was no such method available. The main aim of the present thesis was to introduce a short, structured, age-appropriate interview for young children: the Berkeley Puppet Interview (BPI). In addition, we examined whether information about emotional, behavioral and peer relationship problems, obtained with such an interview, would be valuable for use in epidemiological studies and eventually, clinical settings.

In *Chapter 2* the psychometric properties of eight BPI scales assessing peer relationships and common emotional and behavioral problems were examined. The findings suggested that when children's self-reports are gathered with a structured and developmentally appropriate instrument, as is the BPI, the obtained information is relatively psychometrically sound. Based on confirmatory factor analyses, we demonstrated that the assessed items had a multidimensional structure: a model with more dimensions showed better fit to the data than a model with fewer dimensions. A model with eight factors (Depression, Separation Anxiety, Overanxious, Oppositional Defiant, Overt Hostility, Conduct Problems, Bullied by Peers, and Peer Acceptance/Rejection) had the most favorable fit and fitted equally well in boys and girls. The results also indicated that the BPI seemed a valid measure: the eight BPI scales were associated with known correlates of emotional and behavioral problems, for example higher scores on most BPI scales were associated with lower maternal education, lower family income and non-Western ethnicity. As expected, boys reported more behavioral problems, whereas girls reported more emotional problems. In addition, associations between socio-demographic characteristics and BPI problem scores were largely independent of maternal reports of these problems. The reliability of the coding procedure was high with intraclass correlation coefficients ranging between .87 and .96. Estimates for the reliability (Cronbach's alpha) of each of the eight scales were only modest. Children seemed to provide more reliable information on broad domains of emotional, behavioral and peer relationship problems. Also, correlations between child reported problem behavior and parent reported problem behavior were small. Further, no test–retest reliability was assessed because children visited the research center where the BPI was assessed only once. Despite some caveats, the findings indicate that young children from socioeconomically and demographically diverse backgrounds are capable of

providing valid, multidimensional information on their emotional, behavioral, and peer relationship problems using the BPI.

In *Chapter 3* the validity of the information obtained with the BPI was further examined. The study investigated the extent to which child reported emotional and behavioral problems were related to parent reported DSM-based diagnoses obtained with a standardized diagnostic interview, and to parent reported emotional and behavioral problem scores in the clinical range on DSM-oriented scales of the Child Behavior Checklist (CBCL). Given that agreement between standardized diagnostic interviews and clinical evaluations is mostly modest, it was also examined whether higher BPI problems scores were associated with treatment referral. Results suggested that children with a DSM-based diagnosis obtained by a structured parental interview reported higher levels of emotional and behavioral problems on the BPI. Likewise, children with parent reported CBCL problem scores in the clinical range reported more emotional and behavioral problems. Nevertheless, the reports obtained with the BPI did not predict treatment referral over and above DSM-based diagnoses derived by parent interview. The young children that were assessed thus seemed to have insight in problems that were acknowledged by their parents, but their information had limited value in predicting treatment referral, beyond information about problem behavior that was obtained from the parents.

A common finding in studies of child problem behavior is that agreement between parent and child report of emotional and behavioral problems is only small. Yet, there can be useful information in the diverging reports of informants. The study described in *Chapter 4* examined determinants of discrepancies between mother and child reports of problem behavior. We first identified patterns within the problem behavior reports of mothers and children. Four patterns were identified, differing in problem level, and the direction and magnitude of mother-child discrepancies: one pattern represented no discrepancies (46%), another represented slight discrepancies (30%), and two patterns represented higher problem levels and more discrepancies. In the latter two patterns, either children (11%) or mothers (13%) reported more problems. Child-related characteristics, and in particular cognitive difficulties, were associated with the likelihood that children reported more problems than their mothers. In contrast, adverse family and parenting characteristics were the most salient determinants of the group in which mothers reported more problems than their children. This knowledge about specific child and family characteristics that contribute to mother-child discrepancies can help to interpret diverging informant's reports.

The study described in *Chapter 5* focused on the value of child reported information in a study of the association between parental depression (prenatally or at child age three years) and six-year-old children's problem behavior. Children of depressed parents are at risk of adverse outcomes. However, evidence regarding the specific effect of paternal as compared to maternal depression on young children's well-being is relatively scarce and mainly based on parental reports of child problems. If the same source is used to provide information on both determinant and outcome,

associations may be inflated due to shared-method variance. Therefore, we examined the association between mothers' and fathers' depression and child problem behavior using reports of mothers, fathers and young children themselves. The results obtained with children's reports of problem behavior suggested that maternal and paternal depression similarly affected young children's well-being. However, if parents reported on both their own depression and on child problems, associations seemed inflated. The latter findings were likely influenced by shared-method variance. Future studies examining the effect of parental depression on children's problem behavior would benefit from considering additional information to parents' perspectives. Including child reported information may help to obtain a more accurate estimate of children's problem behavior.

In *Chapter 6* the consequences of maternal and paternal harsh discipline at child age three years on children's emotional and behavioral problems at age six years were explored. Results showed that mild forms of both maternal and paternal harsh discipline were negatively associated with parent and child reported behavioral problems. Findings for parent and child report of behavioral problems were surprisingly consistent, but findings for child and parent-reported emotional problems were less consistent. Importantly, we found that the association between harsh discipline and child reported outcomes was independent of parent reported problem behavior. This finding, that children's reports not only confirmed parental reports, may indicate that the effect of harsh parenting is even larger than what is estimated based on reports of only one informant. If anything, the results suggest that although information from young children should be treated with some caution, the possibility to obtain information from very young children provides opportunities for instances when parents are unavailable or unwilling to serve as informants on their children's emotional or behavioral problems.

Chapter 7 focused on the association of parenting and children's self-regulation with child reported quality of peer relationships. Many studies have shown that parenting and children's self-regulation impact children's peer relationships. However, only few studies examined both factors together and even fewer examined whether it is parenting per se, or the effect of parenting via children's self-regulation abilities that affect children's early peer relationships. Parenting may influence children's social behavior by modeling adequate social behavior and by providing an encouraging environment in which a child can attain self-regulation skills, which in turn can affect children's social behavior. Results of chapter 7 showed that higher levels of parental harsh discipline and lower levels of maternal sensitivity were associated with children's peer aggression. Higher levels of parental harsh discipline were also associated with more peer relationship problems. Importantly, besides a direct association between parenting and child outcomes, parenting seemed to partly exert its effect through associations with children's self-regulation abilities.

Finally, *Chapter 8* summarized the main findings of the thesis and the interpretation and methodological considerations were discussed. Further, implications for research and practice were proposed together with considerations for future studies. It was concluded that the studies

described in this thesis suggest that it is possible to obtain information about emotional and behavioral problems from young children between four and eight years, if obtained with an age-appropriate and standardized method like the BPI. The information children provide appears to be useful for research applications as it is consistent with information obtained by other informants like parents. In addition, children's information may aid in more accurate assessment of the impact of parental factors like depression and harsh parenting on child problem behavior. However, young children's information as obtained with the BPI will not yield a diagnosis and future research on the clinical implications of the BPI is recommended. Nevertheless, if young children's information is used in concert with other measures, or if it is used to test specific hypotheses about child problem behavior, it may advance researchers' understanding of factors relevant to the early development of problem behavior.

S

SAMENVATTING

Wanneer onderzoekers probleemgedrag willen bestuderen, worden zij geconfronteerd met een ingewikkelde situatie. Men wilt vaak constructen onderzoeken die niet direct te observeren zijn, noch onderzocht kunnen worden met objectieve methoden. Daarom zijn er interviews en vragenlijsten ontwikkeld, die geen van allen een volledig objectief beeld van de problematiek kunnen geven, maar waarvan gedacht wordt dat ze wel een accuraat beeld geven van de bestudeerde problemen, zeker wanneer er gebruik gemaakt wordt van informatie uit meerdere bronnen. Een bijzonder ingewikkelde situatie treedt echter op wanneer men informatie wilt verkrijgen over probleemgedrag bij jonge kinderen. Lange tijd werd gedacht dat jonge kinderen zelf geen informatie kunnen geven over hun emoties en gedrag, daarom werd informatie gewoonlijk verkregen via ouders (voornamelijk moeders), leerkrachten en via gedragsobservaties. Echter, jonge kinderen zelf werden niet gebruikt als informant. Er is namelijk slechts een beperkt aantal gevalideerde methoden beschikbaar om bij jonge kinderen op leeftijdsadequate wijze informatie te verzamelen over emotionele- en gedragsproblemen. Bovendien was een dergelijke methode in Nederland in het geheel niet beschikbaar. Een belangrijk doel van dit proefschrift was dan ook om een kort, gestructureerd en leeftijdsadequaat interview voor jonge kinderen te introduceren: het Berkeley Puppet Interview (BPI). Daarnaast werd onderzocht of informatie over emotionele- en gedragsproblemen en over vriendschapsrelaties, verkregen met het BPI, geschikt zou kunnen zijn voor onderzoek en uiteindelijk, voor klinische toepassingen.

In *Hoofdstuk 2* werd een studie beschreven waarin psychometrische eigenschappen van acht BPI schalen bestudeerd werden. De acht bestudeerde BPI schalen beogen emotionele-, gedrags- en (vriendschaps)relatieproblemen te meten die veel voorkomen bij jonge kinderen. De bevindingen in dit hoofdstuk suggereren dat indien zelf-rapportages van jonge kinderen verkregen worden met een gestructureerd en leeftijdsadequaat instrument, zoals het BPI, deze informatie van redelijke psychometrische kwaliteit is. In de studie is een confirmatieve factoranalyse uitgevoerd die laat zien dat het BPI een multidimensionale structuur lijkt te hebben: een model met meerdere dimensies paste beter bij de data dan een model met minder dimensies. De beste modelfit werd gevonden bij een model met acht factoren (depressie, separatie angst, gegeneraliseerde angst, opstandig gedrag, hostiliteit, gedragsproblemen, gepest worden, en afwijzing door leeftijdsgenoten), de factorstructuur was niet anders voor jongens en meisjes. De resultaten van de studie beschreven in hoofdstuk 2 suggereren ook dat het BPI een valide instrument is: de acht BPI schalen waren geassocieerd met zaken waarvan uit de literatuur bekend is dat deze samenhangen met emotionele- en gedragsproblemen, zoals een lager educatieniveau van moeders, een lager familie-inkomen en niet-Westerse etniciteit. Ook rapporteerden jongens, zoals verwacht, meer gedragsproblemen, terwijl meisjes meer emotionele problemen rapporteerden. Daarnaast werd gevonden dat associaties tussen socio-demografische kenmerken en probleemgedrag gerapporteerd tijdens het BPI, vrijwel onafhankelijk waren van probleemgedrag zoals dat door moeders gerapporteerd werd. De betrouwbaarheid van het coderingsproces was hoog. Echter, de Chronbach's alpha

voor ieder van de afzonderlijke schalen was matig. Deze was hoger voor het brede domein van emotionele-, gedrags- en (vriendschaps)relatieproblemen dan voor specifieke probleemgebieden. Ook waren correlaties tussen probleemgedrag zoals dat door kinderen gerapporteerd werd en probleemgedrag dat door ouders gerapporteerd werd klein. Ondanks enkele beperkingen laten de resultaten zien dat jonge kinderen met diverse socio-economische achtergronden in staat zijn om met het BPI valide, multidimensionele, informatie te geven over hun emotionele problemen, gedrags- en (vriendschaps)relatieproblemen.

In *Hoofdstuk 3* werd de validiteit van de BPI data verder onderzocht. De studie beschreven in dit hoofdstuk bestudeerde de mate waarin door kinderen gerapporteerd emotionele- en gedragsproblemen (verkregen met het BPI) gerelateerd waren aan DSM gebaseerde diagnoses (verkregen middels een gestandaardiseerd diagnostisch interview (DISC-YC) met ouders). Tevens onderzochten we de mate waarin probleemgedrag gerapporteerd tijdens het BPI gerelateerd was aan probleemgedrag in de klinische range op de Child Behavior Checklist (CBCL), deze vragenlijst werd ook door ouders ingevuld. Omdat de overeenstemming tussen gestandaardiseerde diagnostische interviews en diagnosen die in de praktijk gemaakt worden vaak laag is, onderzochten we ook of hogere BPI scores geassocieerd waren met verwijzingen voor behandeling van psychische problemen. De resultaten van de studie suggereerden dat kinderen met een DISC-YC diagnose zelf ook meer emotionele- en gedragsproblemen rapporteren dan kinderen zonder een dergelijke diagnose. Eveneens rapporteerden kinderen die volgens ouders CBCL problemen in de klinische range hadden meer emotionele- en gedragsproblemen op het BPI. Echter, na correctie voor DISC-YC diagnosen, voorspelden hogere BPI scores niet of een kind verwezen zou worden voor behandeling van psychische problemen. Op basis van deze studie werd geconcludeerd dat jonge kinderen inzicht lijken te hebben in problemen die hun ouders ook opmerken, maar dat informatie van kinderen zelf slechts beperkte toegevoegde waarde heeft wanneer het aankomt op het voorspellen van verwijzingen voor behandeling van psychische problemen.

Wanneer er een onderzoek uitgevoerd wordt waarin gebruik gemaakt wordt van informatie van meerdere informanten (bijvoorbeeld ouders en kinderen), dan is het een bekend fenomeen dat de overeenstemming tussen informatie van verschillende informaten laag is. Vaak wordt gedacht dat dit problematisch is. Echter, juist in de discrepantie kan ook bruikbare informatie zitten. In de studie beschreven in *Hoofdstuk 4* werden factoren onderzocht die geassocieerd zijn met moeder-kind discrepanties. Allereerst werden patronen geïdentificeerd op basis van de informatie van moeders en hun kinderen. Vier patronen waren te onderscheiden, die varieerden in de hoeveelheid problemen die gerapporteerd werden, en in de richting en de mate van discrepantie tussen informatie van moeders en kinderen. Het eerste patroon werd gekenmerkt door vrijwel geen discrepantie (46%), het tweede door enige discrepantie (30%), het derde en vierde patroon werden beide gekenmerkt door meer problemen en meer discrepantie. In het derde patroon rapporteerden kinderen meer problemen (11%), terwijl in het vierde moeders meer problemen (13%) rapporteerden. Vervolgens werd onderzocht welke factoren samenhingen met moeder-kind discrepantie. Kind-specifieke

factoren, in het bijzonder cognitieve problemen, waren voornamelijk geassocieerd met het patroon waarin kinderen meer problemen rapporteerden dan hun moeders. Familie gerelateerde factoren waren de meest in het oog springende determinanten van het patroon waarin moeders meer problemen rapporteerden dan hun kinderen. Deze kennis over specifieke kind en familie factoren die bijdragen aan moeder-kind discrepanties kan helpen om informatie uit verschillende bronnen beter te interpreteren.

In *Hoofdstuk 5* werd onderzocht wat de waarde van kind rapportage is, wanneer men de associatie tussen depressie van ouders (prenataal of op driejarige leeftijd van het kind) en emotionele- en gedragsproblemen van zesjarige kinderen bestudeert. Het is bekend dat kinderen van depressieve ouders een grotere kans hebben op het ontwikkelen van problemen. Over de specifieke invloed van depressie van moeders in vergelijking tot depressie van vaders is echter veel minder bekend. Bovendien hebben studies die dat onderzocht hebben zich voornamelijk gebaseerd op informatie van de ouders zelf. Wanneer dezelfde bron gebruikt wordt om informatie te verkrijgen over zowel de determinant als de uitkomst, kunnen associaties vertekend zijn als gevolg van zogenaamde 'gedeelde methode variantie'. Om deze redenen werd in de studie beschreven in hoofdstuk 5 gebruik gemaakt van informatie van moeders, vaders en jonge kinderen zelf. De resultaten die we verkregen met zelfrapportages van kinderen als uitkomst, suggereerden dat depressie van vaders en moeders een vergelijkbaar effect hadden op het welbevinden van jonge kinderen. Echter, wanneer ouders zowel over hun eigen depressie als over het probleemgedrag van hun kind rapporteerden, leken associaties sterk uitvergroot. Mogelijk kunnen dergelijke bevinding verklaard worden door de aanwezigheid van gedeelde methode variantie. Toekomstige studies die het effect van ouderlijke depressie op probleemgedrag van kinderen onderzoeken, kunnen gebaat zijn bij het gebruiken van een extra informant naast de ouder zelf. Zelfrapportages van kinderen zouden bijvoorbeeld kunnen helpen om een accurater beeld te krijgen van de daadwerkelijke problemen die kinderen hebben.

In *Hoofdstuk 6* werd een studie beschreven waarin het verband onderzocht werd tussen een harde opvoedingsstijl van ouders bij kinderen van drie jaar en emotionele- en gedragsproblemen op zesjarige leeftijd. Resultaten suggereerden dat milde vormen van een harde opvoedingsstijl van zowel moeders als van vaders negatief geassocieerd waren met probleemgedrag van kinderen. De bevindingen waren verassend consistent wanneer ouder en kind gerapporteerde gedragsproblemen vergeleken werden, maar bevindingen waren minder consistent wanneer ouder en kind gerapporteerde emotionele problemen onderzocht werden. Een van de belangrijkste bevindingen was dat de associatie tussen een harde opvoedingsstijl en kind gerapporteerde problemen onafhankelijk was van ouder gerapporteerde problemen. Deze bevinding, dat rapportages van kinderen meer informatie geven dan alleen het bevestigen van ouder rapportages, kan mogelijk een aanwijzing zijn dat het effect van een harde opvoedingsstijl sterker is dan wat men normaliter schat op basis van informatie van slechts een enkele informant. De resultaten suggereren in ieder geval dat, al moet informatie van jonge kinderen altijd met enige voorzichtigheid behandeld

worden, informatie van zeer jonge kinderen mogelijk bruikbaar is in situaties waarin ouders niet beschikbaar of niet in staat zijn om informatie te geven over probleemgedrag van hun kinderen.

Hoofdstuk 7 focuste op de associatie tussen opvoedingsstijlen, zelfregulatie van kinderen en de door kinderen zelf-gerapporteerde kwaliteit van relaties met leeftijdsgenootjes. Diverse studies hebben laten zien dat opvoeding en zelfregulatie de vriendschapsrelaties van kinderen kunnen beïnvloeden. Echter, er zijn slechts enkele studies die beide factoren samen bestudeerd hebben. Bovendien hebben nog minder studies onderzocht of het specifiek de opvoeding is, of dat het gaat om het effect van opvoeding via zelfregulatie van kinderen dat de vriendschapsrelaties van kinderen beïnvloedt. De resultaten van de studie beschreven in hoofdstuk 7 laten zien dat een harde opvoedingsstijl positief en sensitiviteit van moeders negatief geassocieerd was met agressie jegens leeftijdsgenoten. Een harde opvoedingsstijl was ook positief geassocieerd met problemen in relaties met leeftijdsgenoten (gepest en buitengesloten worden). Een belangrijke observatie was dat er niet alleen een direct verband gevonden werd tussen opvoedingsstijlen en de kwaliteit van relaties met leeftijdsgenootjes, maar dat er ook gevonden werd dat zelfregulerende vaardigheden van kinderen een mediërende rol leken te spelen in dit verband.

Uiteindelijk werden in *Hoofdstuk 8* de belangrijkste bevindingen samengevat en werden deze bevindingen geïnterpreteerd in een breder kader. Ook werden de methodologische overwegingen en implicaties van de studies besproken. In dat hoofdstuk werd geconcludeerd dat het mogelijk lijkt te zijn om informatie over emotionele- en gedragsproblemen te verkrijgen van jonge kinderen tussen vier en acht jaar, als deze informatie verkregen wordt met een leeftijdsadequaat en gestandaardiseerde methode zoals het BPI. Al kan de informatie van jonge kinderen, verkregen met het BPI, niet gebruikt worden om een diagnose te stellen en al moet er meer onderzoek gedaan worden om uitspraken te kunnen doen over de klinische toepassingen van het BPI, de informatie die kinderen geven tijdens het interview lijkt wel bruikbaar te zijn voor onderzoeksdoeleinden. Resultaten gebaseerd op kind-rapportages zijn consistent met resultaten gebaseerd op informatie van andere informanten zoals ouders. Bovendien lijkt informatie van kinderen te helpen om een accurater beeld te krijgen van de invloed van bijvoorbeeld depressie van ouders en een harde opvoedingsstijl op probleemgedrag van kinderen. Indien informatie uit het BPI gebruikt wordt als aanvulling op bestaande methoden, kan dit mogelijk het inzicht in factoren die van belang zijn voor de vroege ontwikkeling van probleemgedrag bevorderen.

S

Addendum

author affiliations
manuscripts and publications
about the author
phd portofolio
dankwoord

AUTHOR AFFILIATIONS

In alphabetical order

Department of Child and Adolescent Psychiatry and Psychology, Erasmus University Medical Center, Rotterdam, the Netherlands
Basten, M., van der Ende, J., Jansen, P.W., Mackenbach, J.D., Rijlaarsdam, J., Ringoot, A.P., So, P., Steenweg-de Graaff, J., Tiemeier, H., Verhulst, F.C., Verlinden, M.

Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands
Hofman, A., Jaddoe, V.W.V., Tiemeier, H.

Department of Pediatrics, Erasmus University Medical Center, Rotterdam, the Netherlands
Jaddoe, V.W.V.

Department of Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands
Tiemeier, H.

Department of Psychology, University of Oregon, Eugene, USA
Measelle, J.R.

Department of Public Health, Erasmus University Medical Center, the Netherlands
Raat, H.

Institute of Psychology, Erasmus University Rotterdam, the Netherlands
Jansen, P.W.

Riagg Rijnmond, the Netherlands
So, P.

School of Pedagogical and Educational Sciences, Erasmus University Rotterdam, the Netherlands
Kok, R.

The Generation R Study group, Erasmus University medical Center Rotterdam, the Netherlands
Basten, M., Jaddoe, V.W.V., Jansen, P.W., Mackenbach, J.D., Rijlaardam, J., Ringoot, A.P., Steenweg-de Graaff, J., Verlinden, M.

MANUSCRIPTS AND PUBLICATIONS

Ringoot, A. P., Jansen, P. W., Steenweg-de Graaff, J., Measelle, J. R., van der Ende, J., Raat, H., Jaddoe, V. W., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2013). Young children's self-reported emotional, behavioral, and peer problems: The Berkeley Puppet Interview. *Psychological Assessment*, *25*(4), 1273-1285.

Stone, L., Daal, C. v., Ringoot, A., Engels, R., Janssens, J., & Otten, R. (2013). Het Berkeley Puppet Interview: Een screeningsinstrument voor psychopathologie bij jonge kinderen. *GZ-psychologie, 5*(2), 18-27.

Verlinden, M., Veenstra, R., Ringoot, A. P., Jansen, P. W., Raat, H., Hofman, A., Jaddoe, V. W., Verhulst, F. C., & Tiemeier, H. (2014). Detecting bullying in early elementary school with a computerized peer-nomination instrument. *Psychological Assessment*, *26*(2), 628-641.

Mackenbach, J. D., Ringoot, A. P., van der Ende, J., Verhulst, F. C., Jaddoe, V. W., Hofman, A., Jansen, P. W., & Tiemeier, H. W. (2014). Exploring the relation of harsh parental discipline with child emotional and behavioral problems by using multiple informants. The Generation R Study. *PloS One*, *9*(8).

Rijlaarsdam, J., Stevens, G. W., Jansen, P. W., Ringoot, A. P., Jaddoe, V. W., Hofman, A., Ayer, L., Verhulst, F. C., Hudziak, J. J., & Tiemeier, H. (2014). Maternal childhood maltreatment and offspring emotional and behavioral problems maternal and paternal mechanisms of risk transmission. *Child Maltreatment*, *19*(2), 67-78.

Ringoot, A. P., van der Ende, J., Jansen, P. W., Measelle, J. R., Basten, M., So, P., Jaddoe, V. W., Hofman, A., Verhulst, F. C., & Tiemeier, H. (2015). Why mothers and young children agree or disagree in their reports of the child's problem behavior. *Child Psychiatry & Human Development*. Advance online publication. doi: 10.1007/s10578-014-0531-x

Thijssen, S., Ringoot, A.P., Wildeboer, A., Bakermans-Kranenburg, M.J., el Marroun, H., Hofman, A., Jaddoe, V.W.V., Verhulst, F., Tiemeier, H., van IJzendoorn, M.H., & White, T. (2015). Brain morphology of childhood aggressive behavior: A multi-informant study in school-age children. *Cognitive, Affective, & Behavioral Neuroscience*. Advance online publication. doi: 10.3758/s13415-015-0344-9

Ringoot, A.P., Tiemeier, H., Jaddoe, V.W.V., So, P., Hofman, H., Verhulst, F.C., & Jansen, P.W., (accepted for publication). Parental depression and child well-being: young children's self-reports helped addressing biases in parent reports. *Journal of Clinical Epidemiology*.

Ringoot, A.P., Jansen, P.W., Rijlaarsdam, J., So, P., Jaddoe, V.W.V., Hofman, A., Verhulst, F.C., & Tiemeier, H. Self-reported problem behavior in young children with and without a DSM-based diagnosis in the general population. *Submitted for publication*.

ABOUT THE AUTHOR

Ank Ringoot was born on September 13th 1984, in Hulst, the Netherlands. After obtaining her gymnasium diploma at the Reynaert College in Hulst in the summer of 2002, she moved to Maastricht to study General Health Sciences at Maastricht University. She completed her Bachelor of Science Degree (2005; cum laude), during which she studied for one semester at the University of Gävle, Sweden. After a clinical internship at the Vincent van Gogh Institute and a research period at the Department of Cognitive Neuroscience of Maastricht University, she obtained her Master of Science Degree in Mental Health Sciences at Maastricht University (2007; cum laude). From 2007 to 2010, she worked as a psychologist at the Vincent van Gogh Institute. In 2010, she started the work described in this thesis at the Generation R Study for the Department of Child and Adolescent Psychiatry and Psychology of the Erasmus Medical Center in Rotterdam. During her PhD, she obtained a Master of Science Degree in Epidemiology from the Netherlands Institute for Health Sciences (2012). Besides the research described in this thesis, she gained clinical experience in child and youth psychology at the outpatient clinic for Child and Youth Psychiatry and Psychology of Sophia Children's Hospital in Rotterdam. In December 2014, she started working as an assistant professor at the Faculty of Psychology and Educational Sciences of the Open University.

PHD PORTFOLIO

Summary of PhD training and teaching activities

Name PhD student:	A.P. Ringoot
Erasmus MC Department:	Kinder- en Jeugdpsychiatrie/psychologie
PhD period:	2010 - 2014
Promotor(s):	Prof. dr. H. Tiemeier, Prof. dr. F.C. Verhulst
Co-promotor:	Dr. P.W. Jansen
Research School:	NIHES

	Year	Workload (ECTS)
1. PhD training		
General courses		
• MSc Epidemiology, Netherlands Institute for Health Sciences (NIHES), Erasmus Medical Center, Rotterdam.	2010-2012	70
• Biomedical English Writing and Communication, Erasmus Medical Center, Rotterdam.	2012	4
Specific courses		
• Mplus course (The Big Mplus Show), Utrecht University.	2011	2
• Teach the Teacher 1, Basic didactic principles, Erasmus Medical Center, Rotterdam.	2013	0.7
• Workshop Individual supervision, Erasmus Medical Center, Rotterdam.	2014	0.1
National and international conferences, seminars and workshops		
• Kick-Off meeting collaboration ErasmusMC KJP and Riagg Rijnmond.	2010	0.3
• Slotsymposium 'GeestKrachtig vooruit' March 24, 2011, Zwolle, the Netherlands. Participation in the 'Information market'.	2011	0.5
• Research Meeting Generation R Oral presentation: The young child as reporter on emotional and behavioural problems – Berkeley Puppet Interview.	2012	0.5
• Life History Society Meeting – London Oral presentation: The young child as informant on emotional and behavioural problems: Findings from 6,428 Berkeley Puppet Interviews in a population-based cohort.	2012	1
• Society for Research in Child Development (SRCD) – Seattle Poster presentation: Determinants of disagreement in maternal and child reports of young children's problem behavior.	2013	1
• Sophia Researchday 2013 – Rotterdam Oral presentation: The Berkeley Puppet Interview – Determinants of discrepant mother-child reports.	2013	0.5

Seminars and workshops

• Attending seminars of the Departments of Psychiatry and Child and Youth Psychiatry/Psychology.	2010-2014	0.5
• Attending the Generation R Study Group research meetings.	2010-2014	1.0
• Symposium: 'Standardized assessment of child psychopathology: new developments'. Erasmus Medical Center, Rotterdam.	2011	0.3
• Symposium: 'Sophia 150 years: children of the future'. Erasmus Medical Center, Rotterdam.	2013	0.3

Other

• Deputy PhD students in the Generation R Management Team.	2011-2012	2

2. Teaching activities

Practicals, workshops and lectures

• Training students and research assistants the coding and interviewing procedure of the Berkeley Puppet Interview, Erasmus Medical Center	2010-2011	3
• Guest lecture 'Assessment of young children. The Berkeley Puppet Interview'. Department of Psychology, Erasmus University Rotterdam.	2011	0.5
• Lecturing VO.3. 'Normal and pathological development of children'. Department of Medicine, Erasmus University Rotterdam.	2011-2014	All 0.5
• Invited training Berkeley Puppet Interview. Behavioural Science Institute, Radboud University Nijmegen.	2012	0.5
• Wetenschapsknooppunt: Teaching scientific research in primary schools	2012	0.5
• Organisation of an information meeting regarding research for students from the Departments of Pedagogy and Psychology, Erasmus University Rotterdam	2012	0.5
• Training students and research assistants in assessing a 'life-events interview', Erasmus Medical Center	2012- 2014	0.5
• Guest lecture 'Assessment of abnormal behavior in young children.' Department of Pedagogics, Erasmus University Rotterdam.	2014	0.3
• Invited training Berkeley Puppet Interview, Babylab, University of Copenhagen, Denmark.	2014	1

Supervising Master's theses/ articles

• Supervising two Colombian students for a short internship	2010	0.5
• Eefje Vermeer, Thesis title: "Macrosomia and emotional and behavioural problems in early childhood".	2011-2012	2
• Aileen Keerveld, Thesis title: "Perceived parenting and emotional and behavioural problems in children".	2011-2012	2
• Joreintje Mackenbach, Article title: "Child reported behavioral effects of maternal and paternal harsh discipline. The Generation R Study".	2011-2012	3
• Natasja Maassen, Thesis title: "Temperamental characteristics and behavioural problems in children"	2012-2013	2
• Roxanne Gal, Thesis titel: "Infant-Mother Attachment and child self-reported problems at age 6".	2013	2
• Rebecca Doualla, Thesis title: "Brain structures associated with self-reported aggression in young children".	2013-2014	2

1 ECTS refers to 28 hours work load