Measuring Up: PERSPECTIVES ON PSYCHOMETRICS AND CONTEXTUAL FACTORS of Early Childhood Problems

Lisanne Lilian Stone
Measuring Up: PERSPECTIVES ON PSYCHOMETRICS AND CONTEXTUAL FACTORS of Early Childhood Problems

Proefschrift

ter verkrijging van de graad van doctor aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus, prof. dr. Th.L.M. Engelen
volgens besluit van het college van decanen
in het openbaar te verdedigen op 5 december
om 14.30 precies

Lisanne Lilian Stone
geboren op 10 mei 1985
te Geldrop
Promotiecommissie

Promotoren
Prof. dr. J.M.A.M. Janssens
Prof. dr. R.C.M.E Engels

Copromotor
Dr. R. Otten

Manuscriptcommissie
Prof. dr. A.H.N. Cillessen
Prof. dr. I. Granic
Em. Prof. dr. J.M.A. Riksen-Waltaven
Prof. dr. P.A.C. Van Lier (Vrije Universiteit Amsterdam)
Prof. dr. T.A. Van Yperen (Rijksuniversiteit Groningen)
Contents

Chapter 1 General Introduction 9

Part I Problem Behaviours in Young Children: Psychometrics and Contextual Factors 37

Chapter 2 Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4- to 12-Year-Olds: A Review 39

Chapter 3 The Parent Version of the Strengths and Difficulties Questionnaire Omega as an Alternative to Alpha and a Test for Measurement Invariance 79

Chapter 4 The Strengths and Difficulties Questionnaire: Psychometric Properties of the Parent and Teacher Version in Children Aged 4-7 95

Chapter 5 Relations Between Internalizing and Externalizing Problems in Early Childhood 125

Chapter 6 Friendship Similarities in Internalizing Problems in Early Childhood 149

Chapter 7 The Co-Development of Parenting Stress and Childhood Internalizing and Externalizing Problems 171

Part II Child Perceptions of Problem Behaviours and Parenting Behaviours 193

Chapter 8 The Berkeley Puppet Interview: A Screening Instrument for Measuring Psychopathology in Young Children 195

Chapter 9 Does Parental Psychological Control Relate to Internalizing and Externalizing Problems in Early Childhood? An Examination Using the Berkeley Puppet Interview 217

Chapter 10 Relations Between Parental and Child Separation Anxiety: The Role of Dependency-Oriented Psychological Control 243

Chapter 11 General Discussion 259

Nederlandse Samenvatting 295

Publications 299

Dankwoord 301

Curriculum Vitae 305
According to the World Health Organization (Friedli, 2009), ‘Mental health is produced socially: the presence or absence of mental health is above all a social indicator and therefore requires social, as well as individual solutions’. Therefore, it seems important to study problem behaviour in relation to contextual factors. Which contextual factors play a role in the development of problem behaviour depends on one’s developmental level and tasks (Erikson, 1968). As young children rely on their parents and are involved in peer-to-peer contact daily, studying the role of these actors may be important when examining problem behaviour. Problem behaviour in this thesis pertains to inwardly directed feelings of distress typified as internalizing problems and outwardly directed feelings of distress typified as externalizing problems. Though, how do we as social scientists know of these problems? As these problems are not directly observable by an objective measure, a hypothetical construct is developed to attempt to describe the phenomenon of ‘problem behaviour’ (Cronbach & Meehl, 1955). This theoretically motivated construct is then operationalized, often by devising items that are aggregated to form a questionnaire or interview. In questionnaires and interviews, the complex reality of problem behaviours is thus reduced to a set of items. Therefore, it is important to investigate whether this method of measuring constructs is reliable (i.e. does the questionnaire produce similar results), and valid (i.e. does the questionnaire measure what it is intended for).

Two main issues are investigated in the current thesis. The first issue entails how a screening instrument for psychopathology performs in terms of psychometrics and what the role is of contextual factors in these parent and teacher reported problem behaviours. The second issue pertains to how an age appropriate instrument for assessing young children’s perceptions of problem behaviour and appraisals of parenting performs in terms of psychometrics and what the role of specific parenting dimensions is in child self-reported problem behaviours. This introductory chapter provides a background of these main issues. Finally, the research methods and datasets that were used in this thesis are introduced and an overview of this thesis is given.

**Problem behaviours in young children: psychometrics and contextual factors**

**Screening for psychopathology by parent and teacher report**
The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is a screening instrument for psychopathology which use as a screening and research tool, a treatment-outcome measure, and a part of clinical assessment is expanding rapidly ever since its development (Hermans, Öry, & Schrijvers, 2005). The origination of this instrument may be explained by looking at the development of several psychological subdisciplines. In the 80’s and 90’s attempts were made to bridge the formerly relatively independent academic and clinical subdisciplines of psychology, resulting in the developmental psychopathology perspective which integrates these sub-
disciplines, with its main goal ‘to understand the processes underlying developmental adaptation or dysfunction’ (Cicchetti, 2006). Traditionally, questionnaires were primarily used by academically oriented psychologists, for example for measuring stages of development (e.g., measuring developmental milestones or Piaget’s stages). Yet, currently, questionnaires form a significant part of clinical psychology, with questionnaires being used frequently to assess problem behaviours in children (Tak, Veerman, & De Wit, 2008). This may be the result of the increased conjunction between academic and clinical psychology. It was within this paradigm that the SDQ was developed by psychiatrist Robert Goodman. Also, one may argue that societal developments in which effectiveness is highly valued may have affected the development of this short screening instrument (Dehue, 2008; Verhaeghe, 2012), as early screening and detection of childhood psychopathology are regarded as important. Accordingly, screening and detection of childhood psychopathology should be swift in order to increase effectiveness.

Evidently, the SDQ is not the only screening questionnaire for child problem behaviours. A host of instruments are available for measuring specific aspects of problem behaviours, such as the Spence Children’s Anxiety Scale regarding anxiety (SCAS; Spence, 1998) or the Children’s Scale of Hostility and Aggression for conduct problems (C-SHARP; Farmer & Aman, 2009). Here, we restrict our overview on instruments that focus on the broad psychopathology spectrum, to ensure comparability to the SDQ. The most important instruments in this overview are the Child Behavior Checklist for parent report and the Teacher Report Form for teacher report developed within the Achenbach System of Empirically Based Assessment suitable for children aged 6-18 (ASEBA/CBCL/TRF; Achenbach & Rescorla, 2001). For children aged 1.5-5 the CBCL/1.5-5 and the Caregiver-TRF are available, which both consist of somewhat less items than the versions for older children (100 instead of 118 items), and some reworded items to match the child’s developmental level. In large part though, these questionnaires are comparable to the versions for older children. As evidenced by more than 4,000 citations of its manual, the CBCL and TRF are used frequently by researchers (Achenbach & Rescorla, 2001). The CBCL and TRF both consist of 118 items divided across eight subscales, measuring problems in the domains anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behavior, and aggressive behavior. The psychometric properties of the CBCL and TRF have been extensively studied and found adequate in both community and clinical samples (see for a review Achenbach et al., 2008). Recently, the ASEBA group developed the Brief Problem Monitor, a questionnaire with items drawn from the CBCL and TRF suitable for children aged 6-18 (BP; Achenbach, McConaughy, Ivanova, & Rescorla, 2011). Again, a parent and teacher report version is available, with the former consisting of 19 items and the latter of 18, both divided across three subscales, problems with internalizing, externalizing and attention. Sufficient indices of reliability and criterion validity have been reported in the BPM manual (Achenbach et al., 2011), although these are the only results published on psychometrics regarding the BPM so far. Besides the ASEBA scales, the Conners Rating Scales are frequently used screening instruments for children aged 6-18, with versions available for parent, Conners Parent Rating Scale Revised and teacher report, Conners Teacher Rating Scale Revised (CPRS; Conners, Sitarenios, Parker, & Epstein, 1998a; CTRS; Conners, Sitarenios, Parker, & Epstein, 1998b). The CPRS consists of 57 items divided across seven subscales: cognitive problems, oppositional, hyperactivity-impulsivity, anxious-shy, perfectionism, social problems, and psychosomatic. The CTRS consists of 38 items divided across six subscales: hyperactivity-impulsivity, perfectionism, inattention/ cognitive problems, social problems, oppositionality and anxious/shy. Psychometric properties of the CPRS and CTRS have been found sufficient in several studies (Conners et al., 1998a; Conners et al., 1998b).

Concluding, the CBCL, TRF, BPM, CPRS, and CTRS provide an alternative to the SDQ. However, the use of the CBCL and TRF is hampered by the many items parents and teachers are asked to fill out. Regarding the BPM, a disadvantage is that little research has been conducted into this instrument. Finally, concerning the CPRS and CTRS research is relatively dated thereby questioning the validity of the available norms, and thus the usability of this instrument.

The SDQ intends to measure both psychosocial problems as well as strengths (for example prosocial behaviour) in children and youths aged 3-16 years through a multi-informant approach. Parents and teachers can report difficulties and strengths among 3-16 year-olds, whereas youths aged 11-16 can report on their difficulties and strengths themselves. The questionnaire consists of 25 items equally divided across five scales measuring emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour. Except for the prosocial scale, the combined scale score reflects total difficulties, indicating the severity and the content of the psychosocial problems. The prosocial scale indicates the amount of prosocial characteristics a child shows (Goodman, 1997). In accordance to the increasing use of the SDQ, the literature on this instrument is growing substantially with many studies investigating psychometric properties of the SDQ. Most research on the SDQ has focused on upper primary school-aged children and youngsters attending secondary school. Psychometric properties of the SDQ in these older children have been found sufficient in community (e.g., Koskelainen, Sournander, & Vauras, 2001) and clinical samples (e.g., Becker, Hagenberg, Roessner, Woerner, & Rothenberger, 2004), but research conducted on younger primary school-aged children shows mixed findings (Edmunds, Garrat, Haines, & Blair, 2005; Goodman, Iervolino, Collishaw, Pickles, & Maughan, 2007; Goodman & Scott, 1999; Hawes & Daddis, 2004; Hill & Hughes, 2007; Perren, Stadelmann, Von Wyl, & Von Kitzing, 2007).
2007; Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006). Thus, it is important to review findings for primary school-aged children in order to draw conclusions about the suitability of the SDQ for younger children (Chapter 2).

**Progress in psychometrics**

In evaluating a questionnaire, several psychometric properties are considered important, depending on the type of questionnaire, its design and its purpose (Evers, Sijtma, Lucassen, & Meijer, 2010; Hunsley & Mash, 2007). While it is often reported that an instrument possesses psychometric properties, for example these properties being adequate or inadequate, an instrument does not have psychometric characteristics per se. Rather, indices of reliability and validity give an indication of the properties for a certain purpose and population (Hunsley & Mash, 2007). This distinction is an important one to make because it prevents making generalizations of psychometric properties of an instrument which are not justified. The following psychometric properties are considered most important according to Evers et al., (2010): reliability, construct and criterion validity. What these properties entail exactly and which of these were evaluated for the SDQ is described next. Regarding reliability, the following subtypes were evaluated for the SDQ: internal reliability, the extent to which items produce similar scores, test-retest reliability, the extent to which a questionnaire yields similar results at different time points, and inter-rater agreement, the consensus between different raters. For validity the following psychometric properties have been evaluated: construct validity, the degree to which the SDQ is similar to other theoretical constructs of child psychopathology, criterion validity, the degree to which SDQ scores predict non-test behaviour or outcome criteria. Subtypes of criterion validity include concurrent validity, the degree to which SDQ scores relate to theoretically similar constructs within time, and predictive validity, the ability of the SDQ to predict scores on another criterion measure in the future. A large body of research exists wherein these properties are evaluated, though not simultaneously, with varying sample sizes and age ranges. Therefore, it is important to meta-analyze the findings reported in the literature such that a more comprehensive picture on the psychometric properties of the SDQ can be painted (Chapter 2).

While the majority of the indicators of reliability and validity suffice, indicators of internal reliability and construct validity have been subject to critique (e.g., Bentler, 2009; Revelle & Zinbarg, 2009; Sijtma, 2009). Specifically, although the use of Cronbach’s alpha is common practice for many social scientists in evaluating internal reliability of a questionnaire (Streiner, 2003), the use of alpha is described as problematic for two psychometric reasons. First, alpha is a lower bound to reliability and in many cases a gross underestimate. Second, alpha is often used as being an indicator of internal consistency leading researchers to conclusions that when alpha is acceptable, items must measure the same underlying construct. However, alpha merely implies the degree of interrelatedness of items, which has little to do with the internal structure of a test (Sijtma, 2009). Recently, several alternatives to alpha such as Revelle’s beta, the greatest lower bound and McDonald’s omega have been proposed and examined on their merits. Of these coefficients, omega seems to provide the most accurate estimate of reliability (Revelle & Zinbarg, 2009).

Second, mostly inadequate factor analytical methods and estimation methods were used in assessing the construct validity of the parent and teacher version of the SDQ. Many studies used Principal Component Analysis (PCA), (Becker et al., 2006; Du, Kou, & Coghill, 2008; Hawes & Daddys, 2004; Kashala, Elgen, Sommerfelt, & Tylleskar, 2005; Muris, Meesters, & Van den Berg, 2003; Smedje, Broman, Hetta, & Von Knorring, 1999). PCA is a technique for constructing components of items without extracting the error part of an item (e.g., Bentler & Kano, 1990) in contrast to factor analysis (FA) that uses a factor model separating a true part and an error part. FA is congruent with classical test theory (Crocker & Algina, 1986). The latent five-factor structure of the SDQ defined by Goodman (1997) was based on descriptive formulations of psychopathology (DSM-IV; American Psychiatric Association, 1994). Hence, confirmatory factor analysis (CFA), instead of exploratory techniques such as PCA or FA, is more suited.

While support for the five-factor structure using CFA is growing substantially (Becker, Woemer, Hasselhorn, Banaschewski, & Rothenberger, 2004; Niclasen, Skovgaard, Andersen, Somhovd, & Obel, 2013; Palmieri & Smith, 2007; Sanne, Torsheim, Heiervang, & Stormark, 2009; Van Leeuwen et al., 2006; Van Roy, Veenstra, & Clench-Aas, 2008), important questions regarding construct validity remain unanswered. Comparing groups is often the focus in developmental psychopathology, underscoring the need of testing measurement invariance. Measurement invariance implies that evidence for construct validity is equal across groups (e.g., Meredith, 1993: Vandenberg & Lance, 2000); and meaningful comparisons can be made. Generally, there is consensus on CFA being the most powerful and versatile approach to test for measurement invariance (Steenkamp & Baumgartner, 1998). In sum, it is important to assess reliability using a different indicator than alpha and to thoroughly analyze the factor structure of the parent and teacher version of the SDQ (Chapters 3 and 4).

**Dutch version of the SDQ**

On a national level, the SDQ is being used frequently in the Netherlands by clinicians, researchers and child healthcare professionals (Hermanns et al., 2005; Treffers, 2007). Research is accumulating on the parent and teacher version for children aged 4-7 (Mieloo et al., 2012; Mieloo, Bevaart, Donker, Van Oort, Raat, & Jansen, 2013). The authors state that the SDQ’s total difficulties scale is reliable and valid and recommend the use of this scale for screening purposes, as the reliabilities of the
subscales have been found too low. However, in one study suboptimal factor analytical techniques were used (Mieloo et al., 2013), compromising the findings regarding validity, and in both studies coefficient omega was not applied, obscuring the findings regarding reliability. Also, the predictive validity of the SDQ has not been extensively studied. Therefore, it is unknown whether SDQ scores are related to a criterion measure over time. Consistent with current bi-directional and cascade effects models on development, it is expected that higher problem behaviours, as measured by the SDQ, are associated with inadequate parenting techniques, higher degrees of parenting stress and lower socio-environmental status over time (e.g., Cicchetti, 2006; Masten & Cicchetti, 2010; Sameroff, 2000). Finally, the available norms for the Dutch version of the SDQ are based on a small and selective sample (Goedhart, Treffers, & Van Widenfelt, 2003; Van Vuuren, Diepenmaat, Reijneveld, & Van der Wal, 2008; Vogels, Crone, Hoekstra, & Reijneveld, 2005). Therefore, research applying modern techniques for investigating internal reliability, test-retest reliability, construct, and criterion validity and for establishing norms for the Dutch version of the parent and teacher version of SDQ is essential (Chapter 4). As the SDQ is primarily a screening instrument, it is most suited to investigate its psychometric properties in a community sample.

**Contextual factors in relation to problem behaviours**

As stated above, problem behaviour is thought to develop in relation to its social context. In ecological systems theory, Bronfenbrenner (1977, 1979) theorized that children’s development may be understood best if studied in close relation to environmental settings. Several levels of social systems are delineated, varying from microsystems which are most proximate to the child, to macrosystems comprising socio-cultural factors. In this thesis, we focus on the microsystem, and specifically on the role of peers and parents. In addition, it may be argued that the problem behaviours themselves form a social context, and as such, contribute to it. The phenomenon of different types of problem behaviour being related to each other is well-known and usually referred to as comorbidity (Achenbach, Howell, Quay, & Conners, 1991; Lilienfeld, 2003; Oland & Shaw, 2005). Thus, before turning to the role of peers and parents it was deemed important to investigate how internalizing and externalizing problems are related, and what the role is of contextual factors in this relation. Although an abundance of research has begun to unravel how and why these problems are associated, there is an ongoing debate on which model best explains the interrelatedness between these problem clusters. First, some authors favor the *directional model*, wherein it is stated that internalizing problems affect externalizing problems, or vice versa (Glaser, 1967; Patterson & Capaldi, 1990). Second, other authors argue that latent constructs influence both problem clusters (Fergusson, Lyskey, & Horwood, 1996; Weiss, Süsser, & Caton, 1998). These constructs can refer to various factors: genetic, biological, psychological and contextual. As such, one could speak of a spurious relation between internalizing and externalizing problems, as this model suggests an underlying factor is responsible for the correlation between internalizing and externalizing problems. This model is referred to as the third variables model (Krueger & Markon, 2006). As there is evidence for both the directional and the third variables model, it is deemed important to simultaneously examine a directional model and a third variable model of childhood psychopathology (cf. Lee & Bukowski, 2012; Mathiesen, Sanson, Stoolmiller, & Karevold, 2009). Therefore, we investigated whether and how internalizing and externalizing problems are related over time and whether third variables may explain these relations, if any (Chapter 5).

**The role of peers**

The microsystem encompasses relations and interactions the child has with immediate surroundings, and is hypothesized to impact the child most strongly (Bronfenbrenner 1977,1979). Friends are one of such microsystem factors that may influence the development of problem behaviours. Friends may serve as cognitive and social resources for children, for example by promoting self-esteem and general self-worth (e.g., Bagwell, Newcomb, & Bukowski, 1998; Hartup, 1996), and most children have mutual friends (Hartup & Stevens, 1999). Therefore, studying friendships of children seems to be important. Also, friends provide children with a context for social skills learning and serve as models for future relationship development (Bukowski, Newcomb, & Hartup, 1996). It is widely thought that formation and maintenance of friendships are to a large part driven by preferences of similarities in appearances, such as age, gender, racial and ethnic background, behaviours, and opinions (i.e., selection processes; Aboud & Mendelson, 1996; Rubin, Lynch, Coplan, Rose-Krasnor, & Booth, 1994). This assumption is referred to in the literature as the ‘homophily hypothesis’ or ‘similarity attraction hypothesis’ (Berndt, 1982; Byrne, 1971; Kendal, 1978). It has been argued that reciprocated friendships may provide a primary context for mutual influence, because children have the greatest opportunity to interact and to share their personal feelings within these relationships (e.g., Rose, 2002). In accordance with this hypothesis, older children and adolescent friends have been found to resemble their reciprocal friends with regard to a host of externalizing behaviours (e.g., Engels, Vitaro. Den Exter Blokland, De Kemp, & Scholte, 2004; Poelen, Engels, Van der Vorst, Scholte, & Vermulst, 2007; Poulin et al., 1997; Prinstein, Boergers, & Spirito, 2001; Prinstein, Meade, & Cohen, 2003; Reitz, Dekovic, Meijer, & Engels, 2006; Vitaro, Tremblay, Kerr, Pagani, & Bukowski, 1997), shyness and prosocial behaviour (Haselager, Hartup, Van Lieshout, & Riksen-Walraven, 1998; Mrug, Hoza, & Bukowski, 2004), and to a lesser extent to internalizing problems such as depression (Brendgen, LaMarche, Wanner, & Vitaro, 2010; Hogue...
parents experience parenting stress to some degree (Crnic & Greenberg, 1990; Deater-Deckard, 2004; Deater-Calkins & Keane, 2007; Van Zaik, Van Zaik, Kerr, & Stattin, 2011), and specifically for young children. As prevalence rates of internalizing problems are highest in adolescence and adulthood (e.g., Cole et al., 2002; Keenan-Miller, Hammen, & Brennan, 2007; Tweeney & Nolen-Hoeksema, 2002), research has focused primarily on adolescent and adult internalizing problems (Zahn-Waxler, Kirmes-Dougan, & Slattery, 2000). However, evidence suggests that the foundation for the development of internalizing problems is laid in childhood (Hankin, Badanes, Abela, & Watamura, 2010; Luby, Si, Belden, Tandon, & Spitznagel, 2009; Zahn-Waxler et al., 2000). So, to understand the dramatic increase in internalizing problems during adolescence, investigating factors associated with childhood internalizing problems is essential (Cicchetti, 2006; Luby, 2010; Ollendick & King, 1994; Zahn-Waxler et al., 2000). Indeed, children with internalizing problems are as likely to have reciprocal best friends as other children (Ladd & Burgess, 1999). In addition, negative peer experiences (e.g., peer rejection and victimization) are important predictors of internalizing problems; likewise, difficulties with friendships are associated with several adjustment problems, such as anxiety and depressed affect (Berndt, 1996; Bukowski et al., 1996; Hodges, Boivin, Vitatoe, & Bukowski, 1999; Ladd & Kochenderfer, 1996; Rubin, Coplan, & Bowker, 2009). While similarities regarding externalizing problems in friendships have been extensively reported and internalizing problems are highly co-morbid with externalizing problems in early childhood (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003), extant previous studies did not control for externalizing problems. Thus, the reported findings may have been attributable to concurrent externalizing problems. Therefore, it is important to investigate whether internalizing problems cluster in friendships while controlling for externalizing problems (Chapter 6).

The role of parents

In addition to peers, parents form a significant part of the microsystem, particularly for young children. While it is a common lay belief that raising children brings happiness, this contention is debated in the literature, with many studies reporting that child rearing puts strains on parents (e.g., Eibach & Mock, 2011; Hansen, 2012). These strains have been found to be highest during early childhood (Berk, 2012; Willford, Calkins, & Keane, 2007). The phenomenon of stress associated with child rearing is termed parenting stress by several scholars (Crnic & Greenberg, 1990; Deater-Deckard, 2004). Parenting stress pertains to daily hassles parents experience in raising their children. Parenting stress is one of the prominent sources of stress, as all parents experience parenting stress to some degree (Crnic & Greenberg, 1990; Hakvoort, Bos, Van Balen, & Hermanns, 2012). Moreover, child psychopathology has been linked to parenting stress in particular (Rodriguez, 2011), and in general to maladaptive parenting techniques and parental functioning (e.g., De Haan, Soenens, Dekovic, & Prinz, 2013; Hughes & Guiltine, 2010; Laird, Pettit, Bates, & Dodge, 2003). Taken together, these findings suggest that child problem behaviour may be important in how parental stresses are maintained. Though, it remains unclear whether problem behaviours may affect parenting stress, and how changes in parenting affect problem behaviours or vice versa, as findings are hampered by a lack of longitudinal studies applying a transactional perspective. The transactional, or developmental psychopathology, perspective (Cicchetti, 2006; Sameroff, 2000) and family systems theory (Minuchin, 1985) propose that processes underlying developmental dysfunction are interrelated dynamically. Specifically, bi-directional parent and child influences have been included in theoretical models explaining psychopathology, which are referred to as parent and child effects models, respectively (e.g., Patterson, 1982; Snyder & Stoolmiller, 2002; see Granic & Patterson, 2006). Therefore, it was deemed important to investigate the interrelated development of parenting stress and children’s problem behaviours, in order to address how these factors affect each other over time (Chapter 7).

Child perceptions of problem behaviours and parenting behaviours

Psychometrics of an age-appropriate self-report instrument

As described, the integration between academic and clinical psychology led to questionnaires on problem behaviours being used increasingly in developmental psychology. However, this holds to a lesser extent for young children as informants. While young children are used frequently in psychological research, by observing them and by using them in experiments, it is relatively infrequent that they are asked directly what their perceptions are of themselves and their social environment. Thus, screening instruments seldomly focus on self-reports of young children. Young children are not always considered reliable informants of their own behaviour (Mutsaers, 2009; Scheerenga & Haslett, 2010). Children’s vocabulary and cognitive development may affect their understanding of questions and interfere with the duration of administration (Arseneault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005). Furthermore, it was often doubted whether young children are capable of self-perception, as this concept is related to cognitive development (Edelbrock, Costello, Dulcan, Kalas, & Calabro-Conover, 1985). Moreover, young children are very sensitive to suggestion, which makes interviewing them a challenge and requires specific interviewing skills (Measeelle, Ablow, Cowan, & Cowan, 1998). Still, already in the 80’s, for instance, Harter (1982) showed that children from the age of eight can meaningfully differentiate between various competence scales.
Measelle et al. (1998) stated that children’s self-perceptions can indeed be reliably measured by using an age-appropriate instrument. In the last few years, children’s self-reports are valued increasingly (Arseneault et al., 2005; Ialongo, Edelsohn, & Keillam, 2001; Luby, Belden, Sullivan, & Spitznagel, 2007), with specific self-report questionnaires being available for children from eight years onwards, such as the Child Depression Inventory (CDI; Kovacs, 2001), Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher, Brent, Chiappetta, Bridge, Monga, & Baugher, 1999), and Perceived Competence Scale for Children (PCSC; Harter, 1982). Also, a computerized questionnaire, the Dominic Interactive, which targets the most common psychopathologies in children, is available for children aged 6-11 (Kuijpers, Otten, Krol, Vermulst, & Engels, 2013; Valla, 2000). However, in practice, there is no screening instrument available in the Netherlands, which uses children younger than six years old as informants for the assessment of their psychopathology. The Berkeley Puppet Interview (BPI; Measelle et al., 1998; Morris et al., 2002) is an interactive interviewing technique, developed in the USA and designed to elicit perceptions of 4.5 to 8-year-olds in an age-appropriate way. In previous studies the BPI has proven to be reliable and valid (Ablow et al., 1999; Arseneault et al., 2005; Luby et al., 2007; Measelle et al., 1998; Morris et al., 2002; Ringoot et al., 2013). However, only one study used a longitudinal design (Measelle et al., 1998) and most studies investigated specific problem clusters of the BPI, mitigating conclusions to be drawn regarding this instrument as a whole. Therefore, it is important to investigate the BPI’s psychometric properties while focusing on the instrument as a whole (Chapter 8). Second, research on stability of its scores is warranted, as only one study reported on stability of these young children’s scores (Measelle et al., 1998). Although the authors report moderate stability for BPI scores, the relatively small sample size of this study impedes drawing firm conclusions regarding stability.

Child perceptions of specific parenting dimensions

The Berkeley Puppet Interview enables investigating children’s appraisals of parenting and its link to children’s self perception of problem behaviour. In order to understand the development of problem behaviours, it might be important to include child perceptions of the factors associated with problems, here referred to as microsystem factors (cf. Bronfenbrenner, 1977; 1979). Among these microsystem factors, parents are thought to be highly influential in child development. Indeed, it has been found that child perceptions of parental control are related more strongly to observations of parenting than parental perceptions of parenting are to observations of parenting in children aged 5 (Sessa, Avenevoli, Steinberg, & Morris, 2001). This suggests that young children have a unique perspective on parenting, a perspective that is now largely overlooked in the massive literature on the association between aversive parenting and problem behaviours in children (Taber, 2010). Therefore, we investigated the association between child-reported problem behaviours and child perceptions of aversive parenting. Before outlining our specific research questions, we provide a short overview of the literature on parenting.

For quite a long time, research on parenting has taken a configurational approach, where parenting styles were defined as combinations of responsiveness (warmth) and demandingness (control) (e.g., Baumrind, 1971; Maccoby & Martin, 1983). Since the 1990s, parenting researchers increasingly adopted a dimensional approach, thereby studying the developmental course, correlates, and antecedents of separate parenting dimensions (e.g., Barber, 1996; Gray & Steinberg, 1999). Along with this shift to a dimensional approach came a more differentiated view on the concept of parental control. Specifically, a distinction was proposed between parental behavioural control and psychological control (Barber, 1996; Steinberg, 1990). Behavioural control was defined as parents’ regulation of children’s behaviour through strategies such as limit-setting and monitoring. In contrast, psychological control was defined as control that intruded upon the child’s psychological world. Such control is characterized by manipulative and pressuring tactics including guilt induction, invalidation of the child’s perspective, and love withdrawal (Barber & Harmon, 2002). Because of its intrusive and pressuring nature, psychological control was hypothesized to represent a threat to children’s emerging sense of self and, as such, to increase the likelihood of maladjustment and internalizing problems in particular (Barber, 1996).

There is robust evidence showing that psychological control is related to internalizing problems in adolescents (Barber & Harmon, 2002). Psychological control has sometimes been portrayed as being independence-stifling in nature. Given that processes of individuation and the development towards independence are highly salient during adolescence (Steinberg, 1990), one might expect that the detrimental effects of psychological control might be most pronounced in adolescence. Scholars have argued that psychological control does not necessarily stifle independence but, instead, frustrates a more basic and universal psychological need for autonomy, that is, the need to feel a sense of volition and psychological freedom (Soenens & Vansteenkiste, 2010). According to self-determination theory (SDT; Deci & Ryan, 2000), this need for autonomy is universal and is not bounded by age. Accordingly, parental psychological control would have deleterious effects on child adjustment irrespective of child age (Soenens & Vansteenkiste, 2010). Indeed, effects of parental psychological control have been reported for elementary-school aged children regarding various indices of psychopathology (Casas et al., 2006; El-Sheikh, Hinnant, Kelly, & Erath, 2010; Joussemet al., 2008; Kuppens, Grietens, Onghena, & Michiels, 2009a; Kuppens, Grietens, Onghena, & Michiels, 2009b; Nelson & Coyne, 2009).
Subsequently, Soenens, Vansteenkiste and Luyten (2010) proposed that maternal difficulties with interpersonal relatedness and closeness may lead to specific controlling parenting tactics, termed dependency oriented psychological control. It is argued that love and care are made contingent on the child’s dependence on the parents. Indeed, dependency oriented psychological control was strongly related to parental anxiety regarding separation in adolescents (Soenens et al., 2010). It is hypothesized that psychological control represent a threat to the child’s emerging sense of self (Barber, 1996), as the child may be unable to develop a stable representation of the mother as a caring person. This unstable representation of the mother may lead to fears of loss of love and abandonment when the child attempts to separate from the parent (Blatt, 2004), potentially leading to difficulties in distancing, interpersonal differentiation, and boundary-formation for the child (Hock & Schritzinger, 1992). In line with the notion that parenting tactics aimed at keeping the child in close proximity are associated with anxiety regarding separation, it was found that intrusive parenting was related to internalizing psychopathology (Kins, Soenens, & Beyers, 2011; Mayseless & Schart, 2009; Soenens et al., 2010; Wood, 2006).

Extant research on psychological control and its association with problem behaviours in childhood suffers from some limitations. In the majority of studies parent reports and observations of psychological control, internalizing, and externalizing problems were used. Parent reports of psychological control, as it represents a clearly maladaptive feature of parenting style, are particularly likely to be affected by social desirability (Bornstein, Hahn, & Haynes, 2011). We assume that children are (more) reliable informants of psychological control because child-reports of other parenting dimensions have been found reliable and valid in childhood (Sessa et al., 2001). Furthermore, internalizing problems in childhood are, by definition, intra-individual experiences (Luby, 2010). Consequently, these symptoms are hard to detect for adult informants, leading to low inter-informant agreement of internalizing problems (Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2005). Moreover, even for observable behaviours often associated with externalizing problems, inter-informant agreement is low (Achenbach et al., 1987; Arseneault et al., 2005; De Los Reyes & Kazdin, 2005). Using child-reports may thus be key in assessing associations of psychological control with internalizing and externalizing problems. Though, we do not consider child-reports as ‘optimal’, as no such ‘optimal’ informant is thought to exist (De Los Reyes, 2011). Hence, it is of paramount importance that the context and the perspectives of different informants are taken into account (De Los Reyes, 2011), and it seems sensible to restrict the informant used to one, in order to capture the perception of the child of both his/her problem behaviours and perceptions of parenting while these are not influenced by informant agreement issues. Although a multi-informant approach is propagated in the literature (Achenbach et al., 1987; Hunsely & Mash, 2007), we argue that a more simple model, using one informant, may be best as a first step when investigating associations between psychological control and problem behaviours. Therefore, we examined the hypothesis that psychological control is related to internalizing problems already in early childhood first using only child-reports, and next using parent reports also (Chapters 9 and 10).

Research Methods

We introduce several analytical techniques that are important to the present thesis.

Meta-analytical techniques. In order to review the literature on the SDQ, correlations reported in different studies should be compared. As the N of these studies varied, correlations will be weighted according their sample size. Correlations are transformed into Fisher’s Z-scores in order to enable the calculation of weighted correlations. The normally distributed Fisher’s Z-scores are weighted according to their sample size minus 3, and a weighted mean Fisher’s Z-score was computed by dividing the sum of the weighted Fisher’s Z-scores by the sum of their weights. The weighted mean Z-score is transformed back to a correlation coefficient r (Field, 2001). This technique will be used in Chapter 2, where we review the literature on psychometric properties of the SDQ.

McDonald’s omega. Traditionally, Cronbach’s alpha is used to examine the internal reliability of a measure, although many statisticians have pointed to the limited usefulness of alpha (Bentler, 2009; Schmitt, 1996; Sijtsma, 2009). Recently, McDonald’s omega as an alternative to alpha has been proposed and seems to provide an accurate estimate of reliability, especially when scale distributions are skewed (Revelle & Zinbarg, 2009). Omega is defined as the percentage of a test that measures a specific factor and is decomposed into a general factor g, factors common to some but not all items f, specific factors unique to each item s and random error e (McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005). Omega is calculated based on the results of factor analysis. In order to take skewness of the data and the ordered categorical nature of the SDQ items into account, we will use omega as a coefficient for internal reliability in Chapter 3, 4, 5, and 7.

Structural Equation Modeling. Structural Equation Modeling (SEM) refers to a statistical modeling technique that combines factor analysis and regression analysis (Byrne, 2011). One of its advantages is the possibility to model constructs as latent variables, thereby accounting for measurement error in the model. The latent variables are inferred from directly measured, or observed, variables. Confirmatory factor
analysis is used to test whether a hypothesized factor structure provides adequate fit to the data in Chapters 3, 4, and 6. Cross-lagged modeling refers to a technique for analyzing longitudinal data, wherein bi-directional associations are specified while controlling for previous associations between variables, which is employed in Chapters 5 and 7.

**Latent Growth Curve modeling.** In order to study change, latent growth curve (LGC) analysis is used (Duncan, Duncan, & Strycker, 2006). Univariate LGC modeling is used to estimate change over time, wherein it is examined whether behaviour increases or decreases, at what level it starts (intercept) and at what pace it develops (slope). Furthermore, inter-individual variability of the initial level of behaviour is examined by assessing the variance of the intercept (i.e., do children differ in their mean behaviour levels). Also, inter-individual variability in the change of behaviour is examined by assessing the variance of the slope (i.e., do children differ in the development of behaviour). Further, by assessing the correlation between the initial level of and change in behaviour it is possible to investigate whether the initial level influences the development of behaviour. Finally, in multivariate LGC it is possible to link multiple growth models to each other, enabling investigation of interrelated development. In Chapter 7, both univariate and multivariate LGC will be applied to examine how parenting stress and internalizing and externalizing problems evolve and co-evolve over time.

**Study characteristics**

Three samples were used to answer our research questions. Characteristics of these samples are outlined in Table 1.

### Table 1: Characteristics of the datasets included in the present thesis

<table>
<thead>
<tr>
<th>Sample</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project Name</strong></td>
<td>Kind in Zicht</td>
<td>Berkeley Puppet Interview</td>
<td>Smoke Free Kids</td>
</tr>
<tr>
<td><strong>Design</strong></td>
<td>Cross-sectional and longitudinal</td>
<td>Cross-sectional and longitudinal</td>
<td>Cross-sectional</td>
</tr>
<tr>
<td><strong>Data waves</strong></td>
<td>3</td>
<td>2</td>
<td>5 (1 data wave used)</td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>Online or paper-and-pencil questionnaire mailed, sociometric data at school</td>
<td>Individual interviews at school</td>
<td>Telephone surveys or paper-and-pencil questionnaires mailed</td>
</tr>
<tr>
<td><strong>Informants</strong></td>
<td>Teachers and mothers of children aged 4-7 at T1; children themselves</td>
<td>Children aged 5-8 at T1</td>
<td>Mothers of children aged 9-11 at T1</td>
</tr>
<tr>
<td><strong>Sample size at T1</strong></td>
<td>N&lt;sub&gt;teacher&lt;/sub&gt; = 2,238 N&lt;sub&gt;parent&lt;/sub&gt; = 1,513 N&lt;sub&gt;children&lt;/sub&gt; = 1,831</td>
<td>N&lt;sub&gt;children&lt;/sub&gt; = 298 N&lt;sub&gt;parent&lt;/sub&gt; = 289</td>
<td>N = 1,478</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td>4, 5, 6, 7</td>
<td>8, 9, 10</td>
<td>3</td>
</tr>
</tbody>
</table>

**Overview of this thesis**

Part I of this thesis focuses on problem behaviour and its psychometric properties. We will start with a review of the literature on reliability and validity of the SDQ in Chapter 2, wherein we examined weighted internal reliability, test-retest reliability, inter-rater agreement, construct, concurrent, and predictive validity. In Chapter 3, we tested the factor structure of the parent version of the SDQ and we examined an alternative indicator of internal reliability, McDonald’s omega. Furthermore, we examined measurement invariance of the SDQ. We extended our study of the psychometric properties of the SDQ by reporting on internal reliability, test-retest reliability, construct, and criterion validity of both the parent and teacher version of the SDQ in Chapter 4. We then focus on internalizing and externalizing problems in relation to the social context wherein children develop. In Chapter 5 we examined how internalizing and externalizing problems are related to each other longitudinally. Furthermore, we examined whether inadequate parenting, parenting stress, maternal health and social preference may explain relations, if any, between...
internalizing and externalizing problems. We examined the role of friendships in Chapter 6, specifically we tested whether internalizing problems cluster in friendships of young children. Next, we will turn to the role of parents. In Chapter 7, we focused on bidirectionality of parenting stress and problem behaviour by examining interrelated development of parenting stress and internalizing and externalizing problems and directionality of these effects.

In part II of this thesis the focus is on child self report and child perceptions of parenting. In Chapter 8, we investigate the psychometric properties of a novel age appropriate instrument for assessing self-perceptions and appraisals of young children. In Chapter 9 this instrument is used to investigate whether psychologically controlling parenting is related to internalizing and externalizing problems beyond traditional parenting dimensions cross-sectionally. We elaborated on this study in Chapter 10 by testing whether a specific type of psychological control mediates the relation between parental psychopathology and internalizing problems in children longitudinally. Ending this thesis, Chapter 11 provides a summary and general discussion of this thesis wherein all findings will be discussed from a broad perspective. Further, limitations of this thesis and implications for research and practice will be discussed.

References


CHAPTER 1 GENERAL INTRODUCTION


CHAPTER 1 GENERAL INTRODUCTION

Sijtsma, K. (2009). On the Like, the Misuse, and Very Limited Usefulness of Cronbach’s Alpha. Psychometrika, 74, 107-120.


Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2003). Cronbach’s α, Revelle’s β, and McDonald’s ω H. Their relations with each other and two alternative conceptualizations of reliability. Psychometrika, 70, 123-133.
Problem Behaviours in Young Children: Psychometrics and Contextual Factors
Chapter 2

Psychometric Properties for the Parent and Teacher Version of the Strengths and Difficulties Questionnaire for 4-12 year-olds: A Review

Published as:
CHAPTER 2

REVIEW OF THE PSYCHOMETRIC PROPERTIES OF THE SDQ

Abstract

Since its development, the Strengths and Difficulties Questionnaire (SDQ) has been widely used in both research and practice. The SDQ screens for positive and negative psychological attributes. This review aims to provide an overview of the psychometric properties of the SDQ for 4-12 year olds. Results from 48 studies (N = 131,223) on reliability and validity of the parent and teacher SDQ are summarized quantitatively and descriptively. Internal consistency, test-retest reliability, and inter-rater agreement are satisfactory for the parent and teacher versions. At subscale level, the reliability of the teacher version seemed stronger compared to the parent version. Concerning validity, 15 out of 18 studies confirmed the five-factor structure. Correlations with other measures of psychopathology as well as the screening ability of the SDQ are sufficient. Concerning validity, 15 out of 18 studies confirmed the five-factor structure. Correlations with other measures of psychopathology as well as the screening ability of the SDQ are sufficient. This review shows that the psychometric properties of the SDQ are strong, particularly for the teacher version. For practice, this implies that the use of the SDQ as a screening instrument should be continued. Longitudinal research studies should investigate predictive validity. For both practice and research, we emphasize the use of a multi-informant approach.

Introduction

Many children’s lives are troubled. Psychosocial childhood problems are common; research has shown that between 3% and 18% of all children suffer from some sort of psychopathology (Bourdon, Goodman, Rae, Simpson, & Koretz, 2005; Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; Egger & Angold, 2003; Ford, Goodman, & Meltzer, 2003; Meltzer, Gatward, Goodman, & Ford, 2003; Zwirs et al., 2007). Behavioral disorders, such as oppositional defiant disorder (ODD), conduct disorder, and attention-deficit/hyperactivity disorder (ADHD), and emotional disorders, such as anxiety and depressive disorders are diagnosed most frequently in children (Canino et al., 2004; Egger & Angold, 2003; Ford et al., 2003).

A substantial discrepancy has been found between the prevalence rates and the number of psychosocial problems being treated in childhood (see for a review Costello, Egger, & Angold, 2005). One of the causes of this divergence may be the stigma (Corrigan, 2004) associated with mental health care or limited access to care (Kataoka, Zang, & Wells, 2002). Another explanation might be that psychosocial problems in the community are often not recognized or diagnosed (Costello et al., 2005). This is worrisome given the fact that problems in young children show relative stability over time (Caspi, Moffitt, Newman, & Silva, 1996) and can potentially escalate or progress into psychiatric disorders. Thus, screening children at an early age for mental health problems and delivering early interventions, which might prevent these childhood problems from developing into more severe psychiatric disorders, is of great importance (Harrington, Rutter, & Fombonne, 1996). Though many instruments are available for screening children, The Child Behavior Check List (CBCL; Achenbach, 1991) has long been viewed as the “gold standard” in assessing childhood problems. Recently, attention for early and quick detection of childhood psychopathology has increased. This has created room for other questionnaires than the CBCL to be used as screening instruments. The launch of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) has enabled researchers and clinicians to increase acceptability in respondents by offering a short and partly positively worded questionnaire (Goodman & Scott, 1999). Whereas the CBCL is a very solid instrument in doing in depth-assessment, the SDQ may be more suitable for screening purposes. The SDQ is thus not a replacement of the CBCL by being the new gold standard, but complements the field of childhood psychological assessment by adding a questionnaire which is shorter and quicker than the CBCL. The CBCL remains very useful though as an in-depth questionnaire. The SDQ has quickly become one of the most utilized screening instruments because it is able to measure both problem behavior and competencies at an early age. In the current study, we reviewed studies examining the psychometric properties of the parent and teacher version of the SDQ. The SDQ is a relatively short, user-friendly screening instrument of psychosocial...
problems for children, worded more positively compared to other common questionnaires. Specifically, the SDQ has relatively few items (25 vs. 118) compared to the Child Behavior Check List (CBCL; Achenbach, 1991). Another advantage of the SDQ is that it is free of charge and available online (www.sdqinfo.com). The SDQ fits the current paradigm in the assessment of psychosocial problems, wherein the focus is expanded to include competencies or strengths in addition to assessing the problems (Carr, 2000; Rhee, Furlong, Turner, & Harari, 2001). The SDQ is based on the Rutter Questionnaires, which were developed in the 1960’s (Rutter, 1967). Goodman updated the items of the Rutter Questionnaires according to the current focus in child psychopathology, for example by adding items to concentration, peer relations, and social competence areas (Goodman, 1994; 1997). The update is based on criteria from the Diagnostic and Statistical Manual of mental disorders, fourth edition (American Psychiatric Association, 1994) and the International Classification of Diseases, tenth edition (World Health Organization, 1992). Additionally, the instrument includes a prosocial scale, which was added to make the assessment more acceptable to respondents. Goodman (1994) devised items of the parent version of the prosocial scale, while the teacher version items were based on the Prosocial Behavior Questionnaire (PBQ; Weir & Duveen, 1981). An impact supplement was added to the SDQ, enabling the informants to report on possible burden and distress (Goodman, 1999).

The SDQ intends to measure both psychosocial problems as well as strengths (for example prosocial behavior) in children and youths aged 3-16 years through a multi-informant approach. Parents and teachers can report difficulties and strengths among 3-16 year-olds, whereas youths aged 11-16 can report on their difficulties and strengths themselves. The questionnaire consists of 25 items equally divided across five scales measuring emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behavior. Except for the prosocial scale, the combined scale score reflects total difficulties, indicating the severity and the content of the psychosocial problems. The prosocial scale indicates the amount of prosocial characteristics a child shows (Goodman, 1997).

The impact supplement comprises of eight questions. The first question asks whether the informant thinks the child has a problem, the remaining questions assess chronicity, distress, social impairment, and burden for others. From these questions, three dimensions can be inferred; perceived difficulties (is there a problem), impact score (distress and social incapacity on the child), and a burden rating (do symptoms impose a burden) (Goodman, 1999).

As the SDQ is translated into over 60 languages, it has been widely used as a screening and research tool, a treatment-outcome measure, and a part of clinical assessment. In accordance to the increasing use of the SDQ, the body of research on the psychometric properties of the instrument is also growing substantially. Therefore, an overview of the results on psychometric properties, reliability and validity would be very useful for researchers and practitioners.

The aim of this review is to review the psychometric properties for the parent and teacher version of the SDQ for children aged 4-12 (primary school-aged children). Most research on the SDQ has focused on upper primary school-aged children and youngsters attending secondary school. Psychometric properties of the SDQ in these older children have been found sufficient in community (e.g., Koskelainen, Sourander, & Vauras, 2001) and clinical samples (e.g., Becker, Hagenberg, Roessner, Woemer, & Rothenberger, 2004), but research conducted on lower primary school-aged children shows mixed findings. Thus, it is important to review findings for primary school-aged children in order to draw conclusions about the suitability of the SDQ for younger children.

Having multiple informants reporting on the SDQ is valuable because psychosocial problems may be highly situational (Achenbach, McConaughy, & Howell, 1987; Goodman, Renfrew, & Mullick, 2000). Thus, the rater’s perception of the situation may influence the ratings. Therefore, we have to investigate whether the psychometric properties of the SDQ in these informants differ and, based on the findings, examine possible implications for the use of the SDQ. Further, the utility of the SDQ is different in clinical versus community populations. In a clinical population, we assume the presence of psychosocial problems. Therefore, the SDQ should inform us about types of psychosocial problems, the duration, and perception of these problems. In a community population of children, we assume the presence of some but not all psychosocial problems, hence, the SDQ should be very sensitive in detecting those children in the community that suffer from (developing) psychosocial problems. The aim of the SDQ is thus slightly different in clinical and community populations. Specifically, we report results on internal consistency, test-retest reliability, and inter-rater agreement. As for validity, the results of construct, concurrent, capacity to discriminate and predictive validity are reported.

Method

Search strategy and selection for identification of studies
The electronic databases PsychINFO, PubMed, and ERIC were searched in March 2010 using the search terms “strengths and difficulties questionnaire,” “validity,” and “reliability.” Neither books nor unpublished articles were retrieved from the references. Abstracts of selected studies were thoroughly read in order to determine whether they were potentially eligible for the inclusion in this review. Inclusion criteria were:

- The target population had to be 4-12 years of age. The age was above the range in 27 out of 48 studies. Of those studies, 3.7 % exceeded the age limit by one
year, 7.4 % by two years, 22.2 % by three years, 25.9 % by four years, 29.6 % by five years, 7.4 % by six years, and 3.7 % by seven years. Still, we included these studies in our review, as the results from younger children in those studies are important for our review. Whenever possible, only the results from primary school-aged children were extracted, and the results from secondary school aged children were omitted.

- Studies had to assess the psychometric properties.
- Studies had to use the parent and/or teacher SDQ version but not self-report.
- Reports had to be available in English.

Eventually, \( k = 48 \) studies were eligible for our review. All studies were published as articles in scientific journals. The publication dates of the 48 articles ranged from 1997 to March 2010. Methodological characteristics of each study are summarized in Table 1. The studies that were selected for this review are indicated with an asterisk in the reference list.

**Strategy for analysis**

The results of internal consistency (the extent to which items produce similar scores) (Cronbach, 1951), test-retest reliability (the extent to which a questionnaire yields similar results at different time points), and inter-rater agreement (the consensus between different raters) enabled us to report the outcomes systematically. In addition, a systematic comparison of the results of construct, concurrent, and capacity to discriminate was feasible. One of the most important assets of a questionnaire, the construct validity, here refers to the degree to which the SDQ is similar to other theoretical constructs of child psychopathology (Campbell & Fiske, 1959). Concurrent validity is defined as the degree to which the SDQ scores relate to a theoretically similar construct, represented in a questionnaire. Capacity to discriminate refers to the ability of the SDQ to distinguish between groups that it should theoretically be able to distinguish between. Predictive validity is defined as the ability of the SDQ to predict scores on another criterion measure. As the method of examining predictive validity differs greatly with respect to research design, the results on predictive validity were not reviewed systematically but descriptively.

Reliability results were reported for each subscale as well as for the impact and total difficulties scales. Correlations were obtained and transformed first into Fisher’s Z-scores in order to enable the calculation of weighted correlations. The normally distributed Fisher’s Z-scores were weighted according to their sample size minus 3 and a weighted mean Fisher’s Z-score was computed by dividing the sum of the weighted Fisher’s Z-scores by the sum of their weights. The weighted mean Z-score was transformed back to a correlation coefficient \( r \) (Field, 2001). Weighted mean correlations were reported separately by type of informant, parent, and teacher. Internal consistency values of \( \alpha = .70 \) and below are generally considered low, values between \( \alpha = .70 \) and \( \alpha = .80 \) acceptable, and values of \( \alpha = .80 \) and above good (Cohen, 1977). Time intervals of test-retest reliability varied between two weeks and six months. Generally, test-retest correlations of \( r = .70 \) and above are considered acceptable. Inter-rater agreement between parents and teachers was reported by subscale and total difficulties scale. No results on the impact scale were reported in the reviewed studies. As a rule of thumb, the meta-analytic mean of inter-rater agreement between parents and teachers \( (r = .27) \) (Achenbach et al., 1987) is used as a benchmark of agreement or data quality (Goodman, 2001). This meta-analytic mean was computed by extracting inter-rater agreement results from 41 studies on the CBCL. As the Achenbach et al. study is known as a landmark paper on inter-rater agreement, the use of .27 as a benchmark seems justified.

Item-level factor loadings were extracted from studies assessing construct validity. Factor loadings were not fully comparable due to the application of different extraction methods (like principal component analysis and principal axis factoring) and rotation methods (orthogonal or oblique) in studies using exploratory factor analysis. The estimation methods were different (maximum likelihood or weighted least squares) in studies using confirmative factor analysis. To gain insight into the quality of the measurement model of the SDQ, loadings were categorized into low (< .40), medium (\( \geq .80 \)), and high (\( \geq .70 \)). Also, weighted mean factor loadings were calculated on item-level.

Concurrent validity was reported mainly as the correlation of SDQ measures with measures of psychopathology like the CBCL or other measures of psychopathology. In the reviewed studies that examined capacity to discriminate, Receiver Operating Characteristic (ROC) analyses were conducted to distinguish between high- and low-risk samples, generating the Area under Curve (AUC). An AUC with a value of 1 shows perfect capacity to discriminate and a value of .5 the absence of capacity to discriminate. Sensitivity (i.e., the proportion of children who are correctly identified by the SDQ as having psychosocial problems) and specificity (i.e., the proportion of children who are correctly identified by the SDQ as not having psychosocial problems) results were extracted and summarized. Again, the results were weighted according to their sample size.

Due to unique research designs in some studies, not all results could be captured in tables. Results from these studies are reported descriptively, as are the results on predictive validity.
Table 1 Summary of studies included in the review

<table>
<thead>
<tr>
<th>Nr</th>
<th>Study</th>
<th>N</th>
<th>Age</th>
<th>Informant</th>
<th>Source population</th>
<th>Country</th>
<th>General study aim</th>
<th>Psychometric properties assessed</th>
<th>Standard comparison diagnosis</th>
<th>Diagnostic domains assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Matsuishi et al. (2008)</td>
<td>2,899</td>
<td>4-12</td>
<td>P</td>
<td>CO</td>
<td>Japan</td>
<td>establish norms, factor structure</td>
<td>internal consistency, construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>2</td>
<td>Perren et al. (2007)*</td>
<td>160</td>
<td>5-6</td>
<td>R, T</td>
<td>CO</td>
<td>Switzerland</td>
<td>predictive value of prosocial behavior</td>
<td>predictive validity</td>
<td>Berkeley Puppet Interview</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>4</td>
<td>Van Leeuwen et al. (2006)</td>
<td>3,169</td>
<td>4-8</td>
<td>R, T</td>
<td>CO</td>
<td>Belgium</td>
<td>psychometric properties Dutch version</td>
<td>internal consistency, inter-rater agreement, construct and concurrent validity</td>
<td>CBCL</td>
<td>prosocial, hyperactivity, internalizing</td>
</tr>
<tr>
<td>5</td>
<td>Samad et al. (2005)</td>
<td>112</td>
<td>4-16</td>
<td>P</td>
<td>CL</td>
<td>Pakistan</td>
<td>validity in Pakistan</td>
<td>capacity to discriminate</td>
<td>ICD-10, Pediatric vs. Psychiatric</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>6</td>
<td>Bourdon et al. (2005)</td>
<td>9,878</td>
<td>4-17</td>
<td>P</td>
<td>CO</td>
<td>U.S.</td>
<td>establish norms U.S., evaluate SDQ</td>
<td>internal consistency</td>
<td>service use</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>7</td>
<td>Dickey &amp; Blumberg (2004)</td>
<td>9,574</td>
<td>4-17</td>
<td>P</td>
<td>CO</td>
<td>U.S.</td>
<td>establish factor structure in U.S.</td>
<td>construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>8</td>
<td>Goodman et al. (2004)</td>
<td>1,028</td>
<td>5-17</td>
<td>R, T, S</td>
<td>CL</td>
<td>U.K.</td>
<td>assess SDQ for improving detecting community psychiatric disorders</td>
<td>capacity to discriminate</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>9</td>
<td>Becker et al. (2004)</td>
<td>543</td>
<td>5-17</td>
<td>R, T</td>
<td>CL</td>
<td>Germany</td>
<td>examine reliability and validity of German SDQ</td>
<td>internal consistency, construct, concurrent and capacity to discriminate</td>
<td>CBCL &amp; clinical diagnoses (ICD-10)</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>10</td>
<td>Hawes &amp; Dadds (2004)</td>
<td>1,359</td>
<td>4-9</td>
<td>R, T</td>
<td>CO</td>
<td>Australia</td>
<td>psychometric properties Australian version</td>
<td>internal consistency, concurrent validity</td>
<td>diagnostic interviews</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>11</td>
<td>Malmberg et al. (2003)</td>
<td>493</td>
<td>5-15</td>
<td>P</td>
<td>CO, CL</td>
<td>Sweden</td>
<td>validity of Swedish SDQ</td>
<td>capacity to discriminate</td>
<td>community vs. Psychiatric</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>12</td>
<td>Glazebrook et al. (2003)</td>
<td>10,745</td>
<td>5-15</td>
<td>P</td>
<td>CO, CL</td>
<td>U.K.</td>
<td>screen for utility of SDQ in pediatric clinics</td>
<td>capacity to discriminate</td>
<td>community vs. Clinic</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>13</td>
<td>Mullick &amp; Goodman (2001)</td>
<td>261</td>
<td>4-16</td>
<td>R, T, S</td>
<td>CO, CL</td>
<td>Bangladesh</td>
<td>suitability of the SDQ for detecting psychiatric problems in Bangladesh</td>
<td>capacity to discriminate</td>
<td>community vs. Clinic</td>
<td>Hyperkinesis, conduct disorder, emotional disorders or any psychiatric disorders</td>
</tr>
<tr>
<td>15</td>
<td>Klasen et al. (2000)</td>
<td>273</td>
<td>4-13</td>
<td>P</td>
<td>CO, CL</td>
<td>Germany</td>
<td>compare SDQ with CBCL in Germany</td>
<td>concurrent validity</td>
<td>CBCL and ICD diagnosis</td>
<td>Total, prosocial</td>
</tr>
</tbody>
</table>
## Table 1 Continued

<table>
<thead>
<tr>
<th>Nr</th>
<th>Study</th>
<th>N</th>
<th>Age</th>
<th>Informant</th>
<th>Source population</th>
<th>Country</th>
<th>General study aim</th>
<th>Psychometric properties assessed</th>
<th>Standard comparison diagnosis</th>
<th>Diagnostic domains assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Goodman et al. (2000)</td>
<td>7,984</td>
<td>5-15</td>
<td>P, T, S</td>
<td>CO</td>
<td>U.K.</td>
<td>SDQ as a means for improving detection of child psychiatric disorders in community</td>
<td>capacity to discriminate</td>
<td>SDQ vs. DAWBA</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>17</td>
<td>Goodman &amp; Scott (1999)</td>
<td>132</td>
<td>4-7</td>
<td>P</td>
<td>CL</td>
<td>U.K.</td>
<td>compare SDQ with CBCL in U.K.</td>
<td>concurrent and capacity to discriminate</td>
<td>CBCL</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>18</td>
<td>Goodman (1999)</td>
<td>699</td>
<td>5-15</td>
<td>P, T</td>
<td>CO, CL</td>
<td>U.K.</td>
<td>validation study into extended SDQ version: impact scale</td>
<td>capacity to discriminate</td>
<td>community vs. Clinic</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>19</td>
<td>Goodman (1997)</td>
<td>403</td>
<td>4-16</td>
<td>P, T</td>
<td>CO, CL</td>
<td>U.K.</td>
<td>compare SDQ with Rutter scales</td>
<td>concurrent and capacity to discriminate</td>
<td>Rutter questionnaires</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>20</td>
<td>Du et al. (2008)</td>
<td>1,965</td>
<td>3-17</td>
<td>P, T, S</td>
<td>CO, CL</td>
<td>China (Shanghai)</td>
<td>assess norms, reliability, validity, factor structure</td>
<td>internal consistency, inter-rater agreement, test retest reliability, construct and capacity to discriminate</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>21</td>
<td>Van Roy et al. (2008)</td>
<td>32,914</td>
<td>10-19</td>
<td>P / proxy</td>
<td>S</td>
<td>Norway</td>
<td>construct validity of five factor structure pre-, early, late adolescence</td>
<td>construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>22</td>
<td>Parkes et al. (2008)</td>
<td>818</td>
<td>8-12</td>
<td>P</td>
<td>CL</td>
<td>Europe (8 regions)</td>
<td>describe psychological symptoms in children with cerebral palsy &amp; assess psychometric properties</td>
<td>internal consistency</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>23</td>
<td>Zwirs et al. (2008)*</td>
<td>2,439</td>
<td>6-10</td>
<td>P, T</td>
<td>CO</td>
<td>Netherlands (ethnically diverse)</td>
<td>develop &amp; validate screening instrument for externalizing disorders</td>
<td>internal consistency, capacity to discriminate</td>
<td>none</td>
<td>Externalizing disorder</td>
</tr>
<tr>
<td>24</td>
<td>Mellor &amp; Stokes (2007)</td>
<td>914</td>
<td>7-17</td>
<td>P, T, S</td>
<td>CO</td>
<td>Australia (ethnically diverse)</td>
<td>factor structure</td>
<td>construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>25</td>
<td>Palmieri &amp; Smith (2007)</td>
<td>733</td>
<td>4-17</td>
<td>P</td>
<td>CL</td>
<td>U.S.</td>
<td>structural validity of P version for custodial grandmothers</td>
<td>internal consistency, construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>26</td>
<td>Hysing et al. (2007)</td>
<td>1,040</td>
<td>7-9</td>
<td>P</td>
<td>CO</td>
<td>Norway</td>
<td>evaluate sensitivity &amp; specificity of SDQ in children with chronic illness</td>
<td>concurrent validity</td>
<td>DAWBA</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>27</td>
<td>Becker et al. (2006)</td>
<td>1,459</td>
<td>6-18</td>
<td>P, T</td>
<td>CO</td>
<td>Europe (10 countries)</td>
<td>examine SDQ P version in several European countries</td>
<td>internal consistency, construct and concurrent validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>28</td>
<td>Sharp et al. (2005)*</td>
<td>659</td>
<td>7-11</td>
<td>P, T</td>
<td>CO</td>
<td>U.K.</td>
<td>establish predictive validity for P and T ratings</td>
<td>predictive validity</td>
<td>Parental help-seeking behavior and worry</td>
<td>Total, prosocial</td>
</tr>
</tbody>
</table>
Table 1 Continued

<table>
<thead>
<tr>
<th>Nr</th>
<th>Study</th>
<th>N</th>
<th>Age</th>
<th>Informant</th>
<th>Source population</th>
<th>Country</th>
<th>General study aim</th>
<th>Psychometric properties assessed</th>
<th>Standard comparison diagnosis</th>
<th>Diagnostic domains assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>Kashala et al. (2005)</td>
<td>1,187</td>
<td>7-9</td>
<td>T</td>
<td>CO</td>
<td>Congo (Kinshasa)</td>
<td>pilot SDQ, investigate mental health and association with school problems, etc.</td>
<td>internal consistency, construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>30</td>
<td>Edmundis et al. (2005)</td>
<td>278</td>
<td>5</td>
<td>P</td>
<td>CO</td>
<td>U.K. (London)</td>
<td>evaluate reliability &amp; validity for SDQ in a health study among young children</td>
<td>internal consistency, concurrent validity</td>
<td>Child Health Questionnaire</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>31</td>
<td>Mellor (2004)</td>
<td>917</td>
<td>7-17</td>
<td>P, T, S</td>
<td>CO</td>
<td>Australia</td>
<td>reliability with younger respondents</td>
<td>internal consistency, inter-rater agreement test retest reliability</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>32</td>
<td>Widenfelt et al. (2003)</td>
<td>1,686</td>
<td>8-16</td>
<td>P, T, S</td>
<td>CO</td>
<td>Netherlands</td>
<td>psychometric properties of Dutch SDQ</td>
<td>internal consistency, inter-rater agreement, concurrent validity</td>
<td>CBCL, YSR</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>33</td>
<td>Muris et al. (2003)</td>
<td>562</td>
<td>9-15</td>
<td>P, S</td>
<td>CO</td>
<td>Netherlands</td>
<td>examine psychometric properties</td>
<td>internal consistency, test retest reliability, concurrent validity</td>
<td>CBCL, CDI, RCMAS, ADHDQ</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>34</td>
<td>Koskelainen et al. (2000)</td>
<td>735</td>
<td>7-16</td>
<td>P, T, S</td>
<td>CO</td>
<td>Finland</td>
<td>evaluate psychometric properties</td>
<td>internal consistency, inter-rater agreement, concurrent validity</td>
<td>CBCL, help-seeking questions</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>35</td>
<td>Goodman et al. (2007)</td>
<td>400</td>
<td>5-7</td>
<td>P</td>
<td>CO</td>
<td>U.K.</td>
<td>check whether minor changes to questionnaire alter psychometrics</td>
<td>capacity to discriminate Rutter questionnaires</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>36</td>
<td>Smedje et al. (1999)</td>
<td>900</td>
<td>6-10</td>
<td>P</td>
<td>CO</td>
<td>Sweden</td>
<td>validation study of Swedish SDQ</td>
<td>internal consistency, construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>37</td>
<td>Kaptein et al. (2008)</td>
<td>967</td>
<td>6-12</td>
<td>P</td>
<td>CO, CL</td>
<td>The Netherlands</td>
<td>assess differences in mental health of ID and non-ID children</td>
<td>internal consistency, capacity to discriminate</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>38</td>
<td>Shojaei et al. (2008)</td>
<td>1,348</td>
<td>6-11</td>
<td>P</td>
<td>CO</td>
<td>France</td>
<td>examine psychometric properties</td>
<td>internal consistency socio-demographic data</td>
<td>Total, prosocial</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Rothenberger et al. (2008)</td>
<td>2,406</td>
<td>7-16</td>
<td>P</td>
<td>CO</td>
<td>Germany</td>
<td>examine psychometric properties of parent SDQ</td>
<td>internal consistency, construct validity</td>
<td>none</td>
<td>Total, prosocial</td>
</tr>
<tr>
<td>40</td>
<td>Goodman et al. (2000)</td>
<td>190</td>
<td>4-16</td>
<td>P, T, S</td>
<td>CL</td>
<td>U.K. and Bangladesh</td>
<td>predict type of disorder from the SDQ</td>
<td>concurrent validity clinical diagnosis</td>
<td>Total, prosocial</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Mathai et al. (2004)</td>
<td>130</td>
<td>4-14</td>
<td>P, T, S</td>
<td>CL</td>
<td>Australia</td>
<td>examine agreement between SDQ and CAMHS</td>
<td>concurrent and capacity to discriminate</td>
<td>community child and adolescent mental health service (CAMHS) Hyperactivity / inattention, conduct and emotional problems</td>
<td></td>
</tr>
</tbody>
</table>
Inter-rater agreement
The results of parent and teacher inter-rater agreement correlations from eight studies by weighted mean correlations and by the range of unweighted correlations are presented in Table 4. The weighted mean correlations varied between .26 and .47. All subscales, except the prosocial scale, had a higher mean than the meta-analytic mean of .27.

Construct validity
A review of the results of the five-factor structure for children aged 4-12 is presented in Table 5. In the parent version, the number of factor loadings was summed across 13 studies. Of these 13 studies, six studies examined also the teacher version. It should be noted that Smedje, Broman, Hetta, and Knorring (1999) and Hawes and Dadds (2004) split their sample into boys and girls, each study generating two sets of factor loadings. Sanne, Torsheim, Heiervang, and Stormark (2009) applied both EFA and CFA, which also generated two sets of factor loadings. Therefore, factor loadings for the parent version summed to 16.
For parent and teacher versions, most items showed satisfactory factor loadings > .40 to ≤ .70. For the parent version, highest loadings were found on the hyperactivity-inattention subscale and lowest on the conduct problems subscale. For teachers, highest loadings were found on the prosocial subscale and lowest on the peer problems scale. However, in 11 out of 14 studies, the results of these factor analyses were obtained by conducting exploratory factor analysis (EFA).

Eight studies applied confirmatory factor analysis, however, only four are presented in Table 5 (Palmieri & Smith, 2007; Van Leeuwen et al. (2006), Van Roy et al. (2008), Sanne et al. 2009) because four out of the total of eight studies did not report factor loadings (Becker, Woerner, Hasselhorn, Banaschewski, & Rothenberger, 2004; Dickey & Blumberg, 2004; Hill & Hughes, 2007; Mellor & Stokes, 2007).

These eight studies are discussed below. Dickey and Blumberg (2004) found support for a three-factor structure of prosocial, externalizing, and internalizing problems. Van Leeuwen et al. (2006) examined a five and a three-factor model in two samples. Support was found for the five-factor model for the parent and teacher version. The three-factor model for the parent and teacher version revealed a worse model fit. The findings of Becker et al. (2004), Van Roy et al. (2008) and Sanne et al. (2009) provided support for the five-factor model for both the parent and teacher version but this factor structure was not found by Mellor and Stokes (2007) and was only marginally adequate in Hill and Hughes’ (2007) study. Palmieri and Smith (2007) confirmed the five-factor structure for custodial grandparents.

Table 2: Weighted mean internal consistency results on the SDQ specified by informant

<table>
<thead>
<tr>
<th>Informant</th>
<th>Parent</th>
<th>Range (Parent)</th>
<th>Teacher</th>
<th>Range (Teacher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosocial Behavior</td>
<td>0.67</td>
<td>0.54 – 0.84</td>
<td>0.82</td>
<td>0.79 – 0.86</td>
</tr>
<tr>
<td>Hyperactivity / inattention</td>
<td>0.76</td>
<td>0.58 – 0.85</td>
<td>0.83</td>
<td>0.66 – 0.89</td>
</tr>
<tr>
<td>Emotional Symptoms</td>
<td>0.66</td>
<td>0.60 – 0.76</td>
<td>0.73</td>
<td>0.63 – 0.80</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td>0.58</td>
<td>0.46 – 0.76</td>
<td>0.70</td>
<td>0.63 – 0.84</td>
</tr>
<tr>
<td>Peer Problems</td>
<td>0.53</td>
<td>0.30 – 0.76</td>
<td>0.63</td>
<td>0.35 – 0.77</td>
</tr>
<tr>
<td>Total Difficulties</td>
<td>0.80</td>
<td>0.53 – 0.84</td>
<td>0.82</td>
<td>0.62 – 0.85</td>
</tr>
<tr>
<td>Impact Scores</td>
<td>0.81</td>
<td>0.69 – 0.87</td>
<td>0.85</td>
<td>-</td>
</tr>
<tr>
<td>N</td>
<td>53,691</td>
<td></td>
<td>21,866</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Weighted mean test-retest correlations on the SDQ specified by informant

<table>
<thead>
<tr>
<th>Informant</th>
<th>Parent</th>
<th>Range (Parent)</th>
<th>Teacher</th>
<th>Range (Teacher)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosocial Behavior</td>
<td>0.65</td>
<td>0.43 – 0.78</td>
<td>0.79</td>
<td>0.50 – 0.84</td>
</tr>
<tr>
<td>Hyperactivity / inattention</td>
<td>0.71</td>
<td>0.48 – 0.85</td>
<td>0.85</td>
<td>0.64 – 0.89</td>
</tr>
<tr>
<td>Emotional Symptoms</td>
<td>0.66</td>
<td>0.47 – 0.82</td>
<td>0.72</td>
<td>0.40 – 0.80</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td>0.66</td>
<td>0.52 – 0.89</td>
<td>0.77</td>
<td>0.58 – 0.86</td>
</tr>
<tr>
<td>Peer Problems</td>
<td>0.66</td>
<td>0.61 – 0.91</td>
<td>0.77</td>
<td>0.58 – 0.82</td>
</tr>
<tr>
<td>Total Difficulties</td>
<td>0.76</td>
<td>0.72 – 0.86</td>
<td>0.84</td>
<td>0.55 – 0.90</td>
</tr>
<tr>
<td>Impact Scores</td>
<td>0.57</td>
<td>-</td>
<td>0.68</td>
<td>-</td>
</tr>
<tr>
<td>N</td>
<td>2,852</td>
<td>-</td>
<td>1,693</td>
<td>-</td>
</tr>
</tbody>
</table>

For parent and teacher versions, most items showed satisfactory factor loadings > 0.40 to ≤ 0.70. For the parent version, highest loadings were found on the hyperactivity-inattention subscale and lowest on the conduct problems subscale. For teachers, highest loadings were found on the prosocial subscale and lowest on the peer problems scale. However, in 11 out of 14 studies, the results of these factor analyses were obtained by conducting exploratory factor analysis (EFA).

Eight studies applied confirmatory factor analysis, however, only four are presented in Table 5 (Palmieri & Smith, 2007; Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006; Van Roy, Veenstra, & Clench-Aas, 2008; Sanne et al. 2009) because four out of the total of eight studies did not report factor loadings (Becker, Woerner, Hasselhorn, Banaschewski, & Rothenberger, 2004; Dickey & Blumberg, 2004; Hill & Hughes, 2007; Mellor & Stokes, 2007).

These eight studies are discussed below. Dickey and Blumberg (2004) found support for a three-factor structure of prosocial, externalizing, and internalizing problems. Van Leeuwen et al. (2006) examined a five and a three-factor model in two samples. Support was found for the five-factor model for the parent and teacher version. The three-factor model for the parent and teacher version revealed a worse model fit. The findings of Becker et al. (2004), Van Roy et al. (2008) and Sanne et al. (2009) provided support for the five-factor model for both the parent and teacher version but this factor structure was not found by Mellor and Stokes (2007) and was only marginally adequate in Hill and Hughes’ (2007) study. Palmieri and Smith (2007) confirmed the five-factor structure for custodial grandparents.
Concurrent validity

Regarding results of concurrent validity, weighted SDQ-CBCL correlations and the range of unweighted correlations are presented in Table 6. The presented correlations do not include all CBCL subscales. In the majority of the reviewed studies, SDQ problem scales correlated with the CBCL subscales that covered similar concepts in general, that is, externalizing, attention problems, internalizing, and social problems. Weighted correlations of .76 for both parent (range of unweighted \( r = .70 - .87 \)) and teacher ratings (range of unweighted \( r = .68 - .87 \)) were found between the SDQ total difficulties and CBCL total scales. At the subscale level, conduct problems, externalizing and hyperactivity, and attention problems correlated sufficiently, while emotional symptoms, internalizing and peer problems, and social problems showed correlations below .70. The SDQ impact scale and CBCL Total scale correlated below .70.

SDQ correlations with measures of general psychopathology. The SDQ has correlated with other measures of general psychopathology. High correlations have been found between SDQ total difficulties and Rutter total deviance scales for parent (\( r = .88 \)) and teacher (\( r = .92 \)) ratings (Goodman, 1997). Another study replicated the correlation between SDQ total difficulties and Rutter total deviance scales for parent ratings (\( r = .76 \)) (Goodman, Iervolino, Collishaw, Pickles, & Maughan, 2007). Somewhat lower correlations were found between the parent rated SDQ and the Chinese version of the parent rated Conner’s Parent Symptom Questionnaire (PSQ; Du, Su, & Li, 1995). SDQ total difficulties and PSQ total score had \( r = .63 \), conduct

### Table 5

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Frequencies of factor loadings on item level of the SDQ specified by informant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequencies</td>
<td>Parent N = 43,274 (13 studies)</td>
</tr>
<tr>
<td></td>
<td>&lt;.40</td>
</tr>
<tr>
<td>Prosocial Behavior</td>
<td></td>
</tr>
<tr>
<td>1 Considerate</td>
<td>0</td>
</tr>
<tr>
<td>4 Shares</td>
<td>1</td>
</tr>
<tr>
<td>9 Caring</td>
<td>0</td>
</tr>
<tr>
<td>17 Kind to kids</td>
<td>1</td>
</tr>
<tr>
<td>20 Helps out</td>
<td>1</td>
</tr>
<tr>
<td>Hyperactivity/inattention</td>
<td></td>
</tr>
<tr>
<td>2 Restless</td>
<td>1</td>
</tr>
<tr>
<td>10 Fidgety</td>
<td>0</td>
</tr>
<tr>
<td>15 Distractible</td>
<td>1</td>
</tr>
<tr>
<td>21 Reflective*</td>
<td>1</td>
</tr>
<tr>
<td>25 Persistent*</td>
<td>1</td>
</tr>
<tr>
<td>Emotional Symptoms</td>
<td></td>
</tr>
<tr>
<td>3 Somatic complaints</td>
<td>4</td>
</tr>
<tr>
<td>8 Worries</td>
<td>0</td>
</tr>
<tr>
<td>13 Unhappy</td>
<td>0</td>
</tr>
<tr>
<td>16 Clingy</td>
<td>0</td>
</tr>
<tr>
<td>24 Fears</td>
<td>0</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td></td>
</tr>
<tr>
<td>5 Tempers</td>
<td>3</td>
</tr>
<tr>
<td>7 Obedient*</td>
<td>3</td>
</tr>
<tr>
<td>12 Fights</td>
<td>1</td>
</tr>
<tr>
<td>18 Lies, cheats</td>
<td>3</td>
</tr>
<tr>
<td>22 Steals</td>
<td>2</td>
</tr>
<tr>
<td>Peer Problems</td>
<td></td>
</tr>
<tr>
<td>6 Solitary</td>
<td>3</td>
</tr>
<tr>
<td>11 Good friend*</td>
<td>5</td>
</tr>
</tbody>
</table>

---

Note: Results on construct validity retrieved from the following studies: Becker et al. (2006), Dickey & Blumberg (2004), Du et al. (2008), Goodman (2001), Hawes & Dadds (2004), Hill & Hughes (2007), Matshuishi et al. (2008), Munis et al. (2003), Kashala et al. (2005), Palmeri & Smith (2007), Sanne et al. (2009), Smedje et al. (1999), Van Leeuwen et al. (2008), & Van Roy et al. (2008). *: Items indicated with an asterisk are reversed items.
problems and conduct problems had \( r = .53 \), hyperactivity-inattention and impulsivity-hyperactivity had \( r = .56 \), hyperactivity-inattention and hyperactivity index score had \( r = .61 \) and hyperactivity-inattention and learning problems had \( r = .58 \) (Du, Kou, & Coghill, 2008). The Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA; Gowers et al., 1999), a clinician based mental health assessment tool, has been correlated with the SDQ total difficulties, resulting in moderate correlations for parent \( r = .38 \) and teacher \( r = .46 \). At the subscale level, correlations between HoNOSCA and the hyperactivity-inattention scales of \( r = .33 \) for parent and \( r = .41 \) for teacher ratings have been reported (Mathai, Anderson, & Bourne, 2002).

SDQ correlations with measures of specific psychopathology. The parent rated SDQ correlated with the clinician rated ADHD-RS-IV (DuPaul, Anastopoulos, Power, Reid, Ikeda, & McGoye, 1998) in that total difficulties and total score had \( r = .50 \). At the subscale level, hyperactivity-inattention and hyperactivity-impulsivity had \( r = .54 \). The SDQ prosocial scale correlated with the parent rated Child Health and Illness Profile-Child Edition (CHIP-CE; Riley, Forrest, Starfield, Rebok, Robertson et al., 2004) on the subscales of resilience \( r = .41 \) and risk avoidance \( r = .40 \) (Becker et al., 2006). The SDQ also correlated with the parent rated ADHQD-P (Schotte & Van der Ploeg, 1998) on total difficulties with total score \( r = .67 \), hyperactivity-inattention with total score \( r = .73 \), and at the subscale level on hyperactivity-inattention with attention-deficit \( r = .65 \), and with hyperactivity \( r = .72 \). Correlations have been found between the parent rated SDQ and the parent rated Child Depression Inventory (CDI-P; Kovacs, 1981) in that total difficulties and total score had \( r = .73 \) and emotional symptoms and total score had \( r = .67 \). The parent rated SDQ correlated with the parent rated Revised Children’s Manifest Anxiety Scale (RCMAS-P; Reynolds & Richmond, 1978) in that difficulties and total anxiety score had \( r = .72 \), and emotional symptoms and total anxiety score had \( r = .73 \) (Muris, Meesters, & Van den Berg, 2003).

Associations of the SDQ with the DAWBA, DSM-IV diagnoses, and risk factors in community samples. An SDQ algorithm was developed in order to predict whether any psychiatric disorder is “unlikely,” “possible,” or “probable” (Goodman et al., 2000). With this algorithm, children with a psychiatric diagnosis, as identified by the Development and Well-Being Assessment (DAWBA; Goodman et al., 2000), were correctly classified as probably having a disorder in 77.3 % of the cases. Using the SDQ algorithm, out of the children who were identified as having hyperactivity or conduct-oppositional or emotional disorder diagnosis according to DAWBA, 91 % were rated as probable for a hyperactivity disorder, 60 % were rated as probable for a conduct-oppositional disorder, and 44 % were rated as probable for an emotional disorder (Hysing, Eigen, Gillberg, Atle Lie, & Lundervold, 2007).

The SDQ algorithm was used in a study to generate diagnoses from SDQ scores. These diagnoses were compared to diagnoses given by independent clinicians or clinical teams based on DSM-IV (1994) criteria. Agreement (expressed in the rank-order correlation tau) between SDQ generated and clinical team diagnoses was found for hyperactivity \( \tau = .44 \), and conduct \( \tau = .56 \) and emotional \( \tau = .39 \) disorders. Reasonable correlations were found between SDQ generated and independent clinician diagnoses for hyperactivity \( \tau = .43 \), and conduct \( \tau = .30 \) and emotional \( \tau = .26 \) disorders (Mathai, Anderson, & Bourne, 2004).

Prevalence of DSM-IV (1994) diagnoses of high (extreme 10 % of sample) versus low risk (90 % of sample) groups based on parent and teacher rated SDQ scores differed. SDQ scores were compared with clinical diagnoses, which were assigned based on the DAWBA. Differences in prevalence between high- and low risk groups showed that all (sub)scales were associated with DSM-IV diagnoses. The odds ratio (OR) for having a psychiatric disorder in the high risk group was 15.7 for parent and 15.2 for teacher rated SDQ’s, across the total difficulties scale and the subscales (Goodman, 2001).

A similar study assessed children with the Diagnostic Interview Schedule for Children, Adolescents, and Parents (DISCAP; Holland & Dadds, 1995) and subsequently assigned DSM-IV diagnoses. Significant differences were found between high and low risks groups on each SDQ subscale and the total difficulties scale, indicating that higher scores are associated with a greater probability of being assigned a DSM-IV diagnosis. The odds ratio for having a psychiatric disorder in the high-risk group was 11.7 based on total difficulties and 14.9 based on the impact scale. In addition, severity of psychosocial problems was rated by clinicians and

<table>
<thead>
<tr>
<th>Conduct problems / Externalizing</th>
<th>Informant</th>
<th>Parent</th>
<th>Range</th>
<th>Teacher</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperactivity / Attention problems</td>
<td>0.69</td>
<td>0.64 – 0.78</td>
<td>0.77</td>
<td>.76 – 0.80</td>
<td></td>
</tr>
<tr>
<td>Emotional symptoms / Internalizing</td>
<td>0.64</td>
<td>0.44 – 0.77</td>
<td>0.58</td>
<td>0.40 – 0.80</td>
<td></td>
</tr>
<tr>
<td>Peer problems / Social Problems</td>
<td>0.52</td>
<td>0.41 – 0.75</td>
<td>0.57</td>
<td>0.48 – 0.71</td>
<td></td>
</tr>
<tr>
<td>Total / Total</td>
<td>0.76</td>
<td>0.70 – 0.87</td>
<td>0.76</td>
<td>0.68 – 0.87</td>
<td></td>
</tr>
<tr>
<td>Impact / Total</td>
<td>0.46</td>
<td>0.44 – 0.51</td>
<td>0.53</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>4,590</td>
<td>-</td>
<td>784</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Note. Results on concurrent validity are retrieved from the following studies: Becker et al. (2004), Janssens & Deboutte (2009), Knasen et al. (2003), Koskelainen et al. (2003), Goodman & Scott (1999), Muris et al. (2003), Syed et al. (2009), Van Leeuwen et al. (2006), & Widenfelt et al. (2003).
correlated with parent rated SDQ scores for total difficulties ($r = .47$) and the impact scale ($r = .57$) (Hawes & Dadds, 2004).

Risk factors such as having contact with a mental health professional or general practitioner (GP), attending special education, or having a desire of using these type of services but not being able to afford them have been shown to be associated with high parent rated SDQ scores. Learning disability, ADHD, declining health, and demographic variables, such as living below the poverty line, living in single-parent or reconstituted families, were significantly associated with high parent rated SDQ scores (Bourdon et al., 2009). For 26 children, parent rated SDQ total difficulties was associated with (consideration of) service use ($OR = 8.7$) (Koskelainen et al., 2000). Parent rated SDQ total difficulties ($r = .16$), emotional symptoms ($r = .15$) and peer problems ($r = .15$) were associated with additional service use in 68 children receiving care in a welfare institution. Further, the need for additional help was predicted by the impact score of parents ($OR = 1.37$) and caregivers ($OR = 1.50$) but not by their total difficulties scores ($OR = 1.07$, $OR = 1.03$) (Janssens & Deboutte, 2009).

**Capacity to discriminate**

In Table 7, weighted AUC values are presented by informant. The combined AUC represents a weighted average of the AUC in each study. The AUC’s were weighted by their standard error. For the subscales, prosocial behavior, and peer problems, the AUC values were just above .5, indicating that, for teacher ratings, the ability of these subscales to distinguish between children with diagnoses, and those without, is just above chance level. For the remaining scales, AUC values are satisfactory.

Two studies could not be incorporated in Table 7 because standard errors or upper bounds were not given. Becker et al. (2004) report AUC’s for the total difficulties (.77, .75), emotional symptoms (.69, .65), conduct problems (.81, .82) and hyperactivity-inattention (.77, .80) scales for the parent and teacher version respectively. So, except for the emotional symptoms scale, the SDQ is adequately able to differentiate between children with and without clinical diagnoses. In a study by Lai, Luk, Leung, Wong, Law and Ho (2009) AUC values were reported for emotional symptoms (.79, .70), conduct problems (.89, .86), hyperactivity-inattention (.86, .85), peer problems (.71, .69), prosocial behavior (.60, .69) and total difficulties (.84, .78), for the parent and teacher version.

Samad, Hollis, Prince, and Goodman (2005) and Malmberg, Rydell, and Smedje (2003) assessed sensitivity and specificity of the parent rated total difficulties and impact scales. The percentages of children identified by the SDQ as having a psychiatric disorder and who did have a disorder (true positives) were 69 % and 82.4 % for total difficulties, and respectively 66 % and 82.7 % for the impact scale (true positives). Children who did not have a psychiatric disorder were correctly identified as such (true negatives) 71 % and 85.4 % of the time by total difficulties, and 86 % and 87.8 % of the time by the impact scale. At the subscale level, sensitivity ranged from 56.6 % to 75 % and specificity from 66 % to 88.1 %. Two other studies assessed sensitivity and specificity by combining parent and teacher reports only for the hyperactivity-inattention and emotional and conduct problems subscales. Goodman et al. (2000) found sensitivity to be 89 %, 81 %, and 90 % respectively on the aforementioned subscales in a London sample and 89 %, 86 %, and 86 % in a Dhaka sample. Reported specificity values were 78 %, 80 %, and 47 % in the London sample and 81 %, 84 %, and 82 % in the Dhaka sample. Mathai et al. (2004) reported sensitivity of 44 % for the hyperactivity-inattention scale, indicating that 44 % of children with ADHD symptoms were correctly identified by the scale as such. Children presenting with emotional symptoms were correctly identified as having emotional symptoms in 36 % of the cases. The scale conduct problems identified 93% of the children showing conduct problems correctly. So, the proportion of true positives that are correctly identified by the SDQ was higher for the conduct problems scale, than it was for the hyperactivity-inattention and emotional symptoms scale.

Goodman, Ford, Simmons, Gattard, and Meltzer (2000) and Goodman, Ford, Corbin, and Meltzer (2004) tested sensitivity in a community and clinical samples. Combined parent (or caregiver) and teacher reports yielded sensitivity of 62.1 % and 82.2 % in detecting any psychiatric disorder, respectively in the community and clinical samples. When only parent report was used, sensitivity dropped to 29.8 % in the community sample and to 51.4 % in the clinical sample. For teacher reports only, sensitivity dropped to 34.5 % and 59.8 % in the community and clinical samples.

### Table 7: Weighted Area Under Curves (by SE) on the SDQ specified by informant

<table>
<thead>
<tr>
<th>Informant</th>
<th>Parent</th>
<th>Range</th>
<th>Teacher</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosocial Behavior</td>
<td>0.71</td>
<td>0.39 – 0.82</td>
<td>0.65</td>
<td>0.64 – 0.67</td>
</tr>
<tr>
<td>Hyperactivity / inattention</td>
<td>0.90</td>
<td>0.76 – 0.97</td>
<td>0.95</td>
<td>0.90 – 0.95</td>
</tr>
<tr>
<td>Emotional Symptoms</td>
<td>0.79</td>
<td>0.69 – 0.85</td>
<td>0.84</td>
<td>0.65 – 0.88</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td>0.92</td>
<td>0.68 – 0.97</td>
<td>0.86</td>
<td>0.82 – 0.87</td>
</tr>
<tr>
<td>Peer Problems</td>
<td>0.71</td>
<td>0.49 – 0.78</td>
<td>0.57</td>
<td>0.45 – 0.69</td>
</tr>
<tr>
<td>Total Difficulties</td>
<td>0.87</td>
<td>0.64 – 0.91</td>
<td>0.83</td>
<td>0.65 – 0.91</td>
</tr>
<tr>
<td>Impact Scores</td>
<td>0.86</td>
<td>0.83 – 0.87</td>
<td>0.88</td>
<td>0.85 – 0.89</td>
</tr>
</tbody>
</table>

Note. Results on capacity to discriminate are retrieved from the following studies: Alyahri & Goodman (2006), Du et al. (2008), Goodman (1997), Goodman & Scott (1999), Kasen et al. (2000), Malmberg et al. (2003), Mullick & Goodman (2001), & Samad et al. (2005).
respectively. Sensitivity for detecting conduct-oppositional, hyperkinetic, ADHD, anxiety, depressive, as well as less common disorders were also assessed. Results were comparable to sensitivity found in detecting any other psychiatric disorder, except for detecting anxiety disorder in the community. Sensitivity was only 45.5% for parent and teacher reports combined and even lower for teacher report only, with a detection rate of 15.9%. Parent report correctly identified anxiety disorders 33.8% of the time, a significant difference to teacher report.

When comparing children with and without intellectual disability (ID), 60.9% with ID were found to have an elevated SDQ score compared to 9.8% of children without ID (Kaptein, Jansen, Vogels, & Reijneveld, 2007). A somewhat similar result was obtained for children with Chronic Illness (CI); 20% of them scored high based on parent rated SDQ total difficulties while 11% of children who did not have CI scored high (Hysing et al., 2007). Children attending pediatric outpatient clinics were more than twice as likely to score in the abnormal SDQ range compared to children from the community (OR = 2.33). The chance of scoring in the abnormal range was even greater for children attending a pediatric clinic for brain disorder (OR = 5.8) compared to community children (Glazebrook, Hollis, Heussler, Goodman, & Coates, 2002).

Goodman (1999) directed special attention to the impact scale of the SDQ. The three concepts of the impact scale, perceived difficulties, impact score, and burden rating, showed a different distribution in community and clinical samples ($\chi^2 = 67.8$), confirming the idea that problems of children in the community sample are not perceived as severe as problems of children in the clinical sample. Lastly, SDQ scores differed according to treatment status. Children currently receiving treatment for psychosocial problems had higher SDQ scores ($M = 15.0$) compared to children not receiving treatment ($M = 8.0$) (Hawes & Dadds, 2004).

**Predictive validity**

Evidence for the predictive validity of the SDQ has been found in three studies. The first focused on the stability of parent ratings, the second on help seeking behaviors, and the third on prosocial behavior. Hawes and Dadds (2004) found that SDQ scores remained relatively stable over a 12-month period for the total difficulties $r = .77$ and impact $r = .63$ scales. For the subscales, comparable correlations were found for hyperactivity-inattention, $r = .77$, prosocial, $r = .64$, conduct, $r = .65$, emotional, $r = .71$ and peer problems $r = .61$.

Sharp, Croudace, Gooddy, and Amtmann (2005) found that, over one year, parent and teacher rated SDQ scores predicted parental help seeking behaviors and worry about the child. Over three time points (6-month intervals), parent rated emotional problems were associated with seeking help from family (OR = 1.09). Parent rated total difficulties at 12 months were associated with worries (OR = 1.06). Emotional problems rated by parents at baseline and six months, predicted worries (OR = .85; OR = 1.33). Teacher rated baseline total difficulties scores were associated with seeking help from a GP (OR = 0.17) and from a friend (OR = 14.88). The rate of change in total difficulties rated by teachers was associated with seeking help from school (OR = 1.13) and GP (OR = 1.25). Teacher rated total difficulties at six months were associated with parental worry (OR = 1.12). Peer problems rated by teachers were associated with parental worry six months later (OR = 1.57).

Perren, Stadelmann, von Wyl, and von Kitzing (2007) examined the role of prosocial behavior in kindergarten longitudinally. In addition to parent and teacher SDQ ratings, children were able to perform as informants regarding their problems by using the Berkeley Puppet Interview (BPI; Measelle, Ablow, Cowan, & Cowan, 1996). Emotional symptoms, conduct problems, and hyperactivity-inattention at age five predicted subsequent emotional symptoms, conduct problems, and hyperactivity-inattention, as rated by multiple informants (i.e., parents, teachers, and children) at age six ($\beta = .530; \beta = .500; \beta = .667$ respectively). The level of prosocial behavior, in combination with the level of emotional symptoms at age five, predicted emotional symptoms at age six. Children showing high levels of prosocial behavior and high levels of emotional problems at age five showed the highest level of emotional symptoms at age six, but children exhibiting high levels of prosocial behavior and low levels of emotional symptoms at age five showed the lowest levels of emotional symptoms at age six.

**Discussion**

The aim of this review was to contribute to a better understanding of the psychometric properties of the SDQ. A total of 48 studies were reviewed. Several indications for research and practice regarding reliability and validity of the SDQ follow from this review.

**Internal consistency**

Results from an impressive number of studies show acceptable internal consistency for the total difficulties and impact scale for both parent and teacher ratings. At the subscale level, we found differences between parent and teacher ratings. Except for hyperactivity-inattention scale, which had an adequate internal consistency, the prosocial scale, emotional, conduct, and peer problems scales showed only moderate internal consistencies for parent ratings. For teacher ratings, the peer problems scale showed a moderate alpha while the remaining scales showed adequate internal consistency. The items of the peer problems scale may not reflect the same construct, as alphas for this scale are lowest for parent and teacher versions. The only item measuring problem behavior is, in our opinion, “picked on or
bullied by other children”. Remaining items seem to reflect loneliness on the one hand (rather solitary, tends to play alone; has at least one good friend) and sociability on the other (generally liked by other children; gets on better with adults than with other children).

An explanation for the difference in internal consistency between parents and teachers is that for parents, the items from the subscales may be less one-dimensional than for teachers, which may refer to a halo-effect for teachers (Abikoff, Courtney, Pelham, & Kopelowicz, 1993; Nisbett & Wilson, 1977). Halo effects occur when one class of behavior influences the perception, and thus the rating, of other behaviors. Specifically, halo-effects have been found to influence ratings of ADHD and ODD (Abikoff et al., 1993; Jackson & King, 2002).

Test-retest reliability

The parent version of the SDQ had lower reliability over time compared to the teacher version, specifically at the subscale level. All parent rated subscales, except the hyperactivity-inattention subscale showed correlations below $r = .70$, whereas teacher subscales were all above $r = .70$. The total difficulties scales for parent and teacher ratings showed good test-retest reliability. Only the impact scale showed to be less reliable over time. The moderate over-time correlation for the impact scale may be due to the time interval of 4 to 6 months that was used in the study assessing the impact scale (Goodman, 2001), in contrast to the time interval of 2 weeks to 6 months used in studies assessing the total difficulties scale (Du et al., 2008; Goodman, 1999; Goodman, 2001; Lai et al., 2009; Mellor, 2004; Muris et al., 2003). The difference in parent versus teacher ratings at the subscale level may be explained in that parents are more prone to detect changes in their child’s mood, as they usually spend more time with their child than their teacher does. This may have caused the correlation to be lower for parent than for teacher ratings.

Inter-rater reliability

Compared to the average inter-rater reliability reported for other measures of child psychopathology, the inter-rater reliability between parent and teacher ratings for total scales and subscales was predominantly better (Achenbach et al., 1987). However, reliability remains modest, which is a well-known phenomenon in psychological assessment. Although inter-rater reliability is valuable to test whether children behave similarly across situations, its use may be less valuable as a psychometric property.

Construct validity

In five studies, the proposed five-factor structure was supported for both parent and teacher versions using confirmatory factor analysis. Recently, support was found for the five-factor model for the parent and teacher version in a very large sample (Sanne et al., 2009). Only one study (Dickey & Blumberg, 2004) found more support for a three-factor structure (internalizing, externalizing, and prosocial behavior) for the parent version. An explanation for the difference in factor structure between the studies of Dickey and Blumberg and Becker et al. (2004), which tested only the parent version using CFA, might be cross-cultural inequivalence (Berry, Poortinga, Segall, & Dasen, 2002). Parents from the U.S. may perceive problems differently than German parents do, which could lead to inconsistencies in factor structures.

In this review, most evidence was thus found for the original five-factor structure of prosocial behavior, hyperactivity/inattention, conduct, emotional, and peer problems. An important methodological aspect of construct validity needs to be highlighted. Despite the theoretical foundation for a five-factor structure, non-normal distribution of scores, and a three-item response category, most studies reported results of exploratory factor analysis and principal component analysis. Both techniques are not suited to test the underlying structure of the SDQ. As the SDQ is based on theoretical constructs concerning child psychopathology (Goodman, 1997), scores are non-normally distributed and the response category is limited; therefore, confirmatory factor analysis (CFA) should be the first method of choice when investigating factor structure (Sanne et al., 2009).

Concurrent validity

Many studies, comparable in some but not all cases, have validated the SDQ. Summarizing and interpreting the results from these studies is therefore complex. Correlations between SDQ and CBCL scales showed to be high for both parent and teacher ratings at the total scales level. The SDQ is thought to measure the same constructs as the CBCL, and these high correlations support that notion. However, at the subscale level, evidence for concurrent validity is less clear. The SDQ emotional and peer problems scales correlated moderately with the CBCL internalizing and social problems scales for both parent and teacher ratings. Further inspection of the CBCL internalizing subscales showed that the CBCL Anxious/Depressed subscale is very well represented by providing three out of five items which are very comparable to the items from the SDQ Emotional Symptoms subscale. However, no items from the CBCL Withdrawn subscale and only one from the Somatic Complaints and Emotionally Reactive subscales are represented in the SDQ Emotional Symptoms subscale. The Withdrawn subscale consists of items that reflect the autism spectrum disorders (ASD), which are not included in the SDQ. The overlap between the CBCL internalizing subscales and the SDQ Emotional Symptoms scale is thus quite small, which may explain the moderate correlation found in our review.

The SDQ impact correlated moderately with the CBCL total problems scale for both parent and teacher ratings. Experience of social impairment and substantial
distress caused by psychiatric symptoms is nowadays a part of the diagnostic criteria for a psychiatric disorder (American Psychiatric Association, 1994, World Health Organization, 1992). The CBCL does not contain social impairment and distress items that would be similar to the SDQ impact supplement. Hence, the moderate correlation between the SDQ impact and CBCL total problems scales may indicate that these scales are conceptually different. The impact scale also correlated with a parental burden scale resulting in $r = .74$ (Goodman, 1999). This parental burden scale is thought to be more comparable to the impact scale than is the CBCL total scale. The CBCL total scale focuses on symptoms of psychosocial problems whereas the impact and parental burden scale focuses on the perception of the consequences of psychosocial symptoms.

In addition to the CBCL, the SDQ had a moderate to high correlation with measures of general and specific psychopathology. High correlations were found specifically for the Rutter scales, on which the SDQ is partly based (Goodman, 1997), and for measures of depression and anxiety. This is contradictory to the low correlation found between the SDQ emotional and peer problems scales and the CBCL internalizing and social problems scales. However, as the SDQ correlated with specific measures of depression and anxiety here, the overlap between symptoms may have become greater and thus the correlations higher. Further, in community samples, SDQ scores also detect psychiatric diagnoses assigned by clinicians. Risk factors for developing psychosocial problems, such as poor health, seem to be associated with higher SDQ scores. This indicates that concurrent validity of the SDQ in comparison to different measures of psychopathology, psychiatric diagnoses, and risk factors is well established.

### Capacity to discriminate

The SDQ proves to be a good screening instrument, with high sensitivity and specificity for the total difficulties and impact scales. The percentage of children correctly identified by the SDQ as having a disorder is high, as is the percentage of children correctly identified by the SDQ as not having a disorder. A more detailed insight into the ability of the subscales to distinguish between community and clinical samples is reflected in the AUC values. Weighted AUC values indicate that, for teacher ratings only, the prosocial behavior and peer problems subscales distinguish between children with diagnoses, and those without, at the chance level. Prosocial behavior does not reflect child psychopathology, so it is not expected to distinguish between community and clinical samples. The peer problems scale again showed some inadequacy here.

However, we cannot infer from the sensitivity and specificity values which proportion of children with abnormal test results are truly abnormal (Altman & Bland, 1994). When using the SDQ we should therefore always consider the context, i.e. clinical versus community samples. If used in a community sample, quite a few children with clinical range SDQ results will actually be typically developing, i.e. false positives, due to low prevalence rates in the general population. In contrast, when the SDQ is used in a clinical sample, where prevalence rates are higher, fewer children will be false positives, but more will be false negatives. It is thus important to consider that the accuracy of the SDQ as a screening instrument varies accordingly with the prevalence rates in a certain population. This underscores the need for using multiple diagnostic instruments in clinical or at risk settings, such as pediatric clinics.

### Predictive validity

Only three studies assessed the predictive validity using a longitudinal design. The results showed evidence of predictive validity, as SDQ scores predicted help seeking for psychosocial problems over a year. Two studies found evidence for SDQ scores predicting similar SDQ scores over a year. In addition, they clarified the role of prosocial behavior in the development of psychosocial problems. Prosocial behavior has not been found to be compatible with high levels of internalizing behavior and thus is not beneficial to children showing highly internalizing behaviors, which concurs with the literature (Hay, 1994).

### Conclusion

Overall, the 25-item SDQ shows strong psychometric properties. Shorter scales are usually less reliable compared to longer scales, which means they also tend to attenuate the validity (Streiner & Norman, 1989). However, the SDQ’s brevity did not substantially influence its psychometric properties. As for reliability, internal consistency of the total scales was satisfactory. Ratings showed sufficient reliability over time and agreement between parents and teachers was relatively high. We should note here that these conclusions are stronger for teachers. Results concerning validity are less straightforward, but in general, we may state that the five-factor structure was confirmed by most studies, correlations with other measures of child psychopathology were high, and evidence for the screening ability of the SDQ was convincing. Predictive validity has not been studied extensively yet, so these findings should be interpreted with caution.

Additional attention should be directed to the necessity to conduct longitudinal studies that would examine the predictive validity of the SDQ and to the validation of the prosocial scale. Overall, the peer problems scale showed the weakest reliability and validity results that were most salient for parent ratings. The prosocial scale also showed some weaknesses, especially concerning internal consistency and capacity to discriminate. This notion should be familiar to researchers as these findings on the
superior sociometric status (e.g. has ability to decode peers' feelings), and thus the comparable SDQ prosocial scale (e.g. considerate of other people's feelings). Behavior that is more difficult to observe may be more susceptible to inferences from raters, for example according to the relationship of the rater with the child (e.g. Ladd & Proﬁlet, 1996). Inferences may be stronger for parents than for teachers in rating prosocial behavior, as internal consistency is lower for the former raters. This may be explained by the nature of the relationship with the child which differs clearly for parents versus teachers.

The peer problems scale showed low internal consistency values for both parent and teacher ratings. Peer problems are most often assessed via reports by children themselves, (i.e. sociometrics) because children are regarded “insiders” whereas parents and teachers are regarded “outsiders” of the peer group. Judgments of peers are based on many and varied social interactions with those being assessed, which may be unknown to “outsiders” (Rubin, Coplan, Chen, Buskirk, & Wojslawowicz, 2005). Assessment of peer problems by parents and teachers is further impeded by the adult perspective used to interpret children’s social interactions, the relationship with the child and child's gender (Ladd & Mars, 1986; Ladd & Proﬁlet, 1996; Rubin & Coplan, 1992). The outsider view combined with the mentioned rater biases may be responsible for the low internal consistency values for the peer problems scale found in our review.

Further, parents and teachers observe children in differing contexts, where different behaviors are shown. This may lead to lower values of internal consistency for both the peer problems and the prosocial subscale. As for rater bias, regardless of rater bias being a factor in the weak performance of subscales, it is important to be aware of rater bias when dealing with screening instruments. The application of screening instruments like the SDQ can be meaningful, in the sense that children are screened before psychosocial problems exacerbate, only if they are used appropriately.

Finally, it is important to note that results from this review are only applicable to the parent and teacher version of the SDQ. The SDQ self-report version was not included in this review because it was not developed nor intended to be used for children younger than 12 years of age. From a developmental perspective the use of traditional self-report questionnaires in children younger than 12 years of age has been questioned, and in children younger than 8 years discouraged. Due to limited linguistic, cognitive and social-emotional abilities children were not thought to provide reliable self-reports (Edelbrock, Costello, Dulcan, Kalas, & Calabro-Conover, 1985; Fallon, & Schwab-Stone, 1994).

Recently, tests of using a puppet interview and computerized pictorial questionnaire have yielded results which point to promising psychometric results in children as young as 5-7 and 6-11 years (Measelle et al., 1998; Valla, Bergeron, Bérubé, Gaudet, & St-Georges, 1994). However, the SDQ and the former interview methods differ greatly in respect to taking into account the developmental level of the elementary school child. The former interviews take into account the developmental level of the child by giving both visual (graphics) and auditory stimuli. The cognitive abilities of children below age 12 may not be sufﬁciently developed to adequately respond to the SDQ questions, which are presented only by visual verbal information (Edelbrock et al., 1985; Fallon, & Schwab-Stone, 1994). Therefore, we have focused on the parent and teacher version of the SDQ in this review.

Limitations
Some limitations of this review should be noted. First, the methodologies varied across the reviewed studies, making it sometimes impossible to extract data from those studies. Comparing these studies to each other was therefore difficult and conducting a meta-analysis on the data was not possible. Second, many studies did not state which parent was used as a rater, making it hard to draw speciﬁc conclusions concerning rater bias. In addition, it was beyond our scope to consider rater psycho-pathology. Third, few studies were conducted using a longitudinal design, making it hard to draw robust conclusions regarding predictive validity. In addition, the reviewed studies did not give sufﬁcient attention to validation of the prosocial scale. Future research should reveal whether the SDQ predicts psychosocial problems and whether the prosocial scale correlates with other measures of prosocial behavior.

Implications
With these limitations in mind, the implications of these results for practice and research can be noted. This review offers researchers and clinicians a clear overview of the psychometric properties for the parent and teacher versions of the SDQ for
4-12 year-olds. Reliability and validity results at the subscale level have been found weaker as compared to the results for the total scales. Therefore, caution is warranted when using and interpreting the subscales of the SDQ separately. Sanne et al. (2009) argued that the distinctiveness of the subscales is not convincing. An explanation for this may be the high comorbidity of psychosocial problems (Ford et al., 2003). Moreover, caution is warranted if a single informant reports on the SDQ, as results may not generalize to other contexts. The use of multiple informants should always be priority when using the SDQ. Most studies used parents and teachers, but possibilities of using other informants should be explored. For example, neighbors, daycare workers, or sports club coaches might be able to report on children’s psychosocial problems. Future research should reveal whether these informants are able to assess psychosocial problems reliably.

For clinical practice in particular, the SDQ is a useful instrument for quickly assessing possible psychosocial problems. The results found in this review give rise to some specific implications at the subscale level. First, the prosocial subscale shows some weaknesses in its psychometric properties, especially for the parent version. Low levels of prosocial behavior and high levels of aggression have been shown to increase the risk for future social adjustment difficulties (Coie, Dodge, & Coppotelli, 1982; Crick, 1996; Romanos, Tremblay, Boulerice, & Swisher, 2005). Excessively high levels of prosocial behavior are also a risk factor for psychopathology (Hay, 1994; Perren et al., 2007), underscoring the importance of assessing prosocial behavior. Therefore, when assessing prosocial behavior teacher ratings should always be included in addition to parent ratings. Further assessment of the child, for example by observing the child in the class room or a naturalistic play situation, should reveal whether the reported lack of prosocial behavior is confirmed by a mental health specialist. When a child is referred for treatment, interventions target at the increase of prosocial behavior instead of the decrease of aversive behaviors (Coie & Koppel, 1990). This emphasizes the importance of assessing prosocial behavior adequately.

Second, the psychometric properties of the hyperactivity/inattention scale are adequate and the SDQ should thus provide a reliable and valid report as to whether ADHD symptoms are present. However, when an ADHD diagnosis is suspected, identification of one of the subtypes Inattentive, Hyperactive-Impulsive or Combined is required (American Psychiatric Association, 1994). Further assessment may be done by using one of the many ADHD rating scales available, such as the SNAP-IV (Swanson, 1992) or the SWAN (Swanson et al., 2005).

For the emotional symptoms subscale, psychometric properties are also adequate. However, in contrast to externalizing problems, internalizing problems are reported more accurately by children themselves than by their parents and teachers (Edelbrock et al., 1985; Ederer, 2004). Gaining insight in the child’s subjective experience of its emotional symptoms is thus highly relevant and advisable in clinical settings.

The conduct problems subscale shows adequate reliability and validity. In order to assess whether a diagnosis of Oppositional Defiant Disorder or Conduct Disorder would be justified, additional assessment is indicated. Because children themselves tend to underestimate their externalizing problems, parents and teachers are particularly important in the further assessment of children presenting with conduct problems (Loeber, Green, Lahey, & Stouthamer-Loeber, 1991).

Finally, the psychometric properties of the peer problems scale are quite weak in some respects. Assessing peer problems is complicated because children are considered as “insiders” who contribute unique information about their peer group. Possibly, it is difficult for parents and teachers to estimate the problems children experience in their peer group because they are “outsiders”. Because of the difficulties with assessing peer problems, additional assessment is essential. The Perceived Competence Scale for Children (Harter, 1982) is a very suitable measure for this purpose. Further, classroom observation is recommended (Wragg, 1994).

The SDQ is not intended to be used as a psychiatric diagnostic instrument; therefore, should not be utilized as such. As a screening instrument, the SDQ performs very well and adds to the field of early detection of child psychopathology. The SDQ has been translated into over 60 languages, which is a great benefit. However, norms are available only for six countries. Culture plays a role in the distribution and expression of psychosocial problems in society, and thus norms for every culture should be established. Results from studies assessing capacity to discriminate showed that the SDQ distinguishes well between children with and those without diagnoses. In populations at risk for psychosocial problems, such as children attending pediatric clinics, we recommend screening of all children referred to specialist services.

For research purposes, longitudinal designs should be employed in order to assess predictive validity more thoroughly. The SDQ is a promising instrument for researching developmental pathways, as it seems to be well validated, short, and acceptable. The teacher version shows strong psychometric properties, but our review shows that the parent version is at the focus of research (17 out of 48 studies studied only the parent version of the SDQ). However, researchers do not fully employ the use of a multiple informant approach. We do argue for such a multi-informant approach, as it is essential for children, their parents, and society when psychosocial problems are found at a young age.

Acknowledgements
The authors would like to thank Liesbeth Linssen for her advice regarding statistics.
References


Chapter 3

The Parent Version of the Strengths and Difficulties Questionnaire: Omega as an Alternative to Alpha and a Test for Measurement Invariance

Published as:
CHAPTER 3 OMEGA AS ALTERNATIVE TO ALPHA

Abstract

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is a widely used screening instrument in mental health care and research. The SDQ has become a favorable instrument because of its brevity, free availability and multi-informant approach. A number of studies have provided insight into its psychometric properties (e.g. Goodman, 2001; Rothenberger, Becker, Erhart, Wille, & Ravens-Sieberer, 2008; Smedje, Bromlan, Hetta, & von Knorring, 1999), which are often described as acceptable; however a recent review showed that reliability (i.e. Cronbach’s alpha) seems insufficient for the conduct and peer problems subscales (Stone, Otten, Engels, Vermulst, & Janssens, 2010). The reliability and factor structure of an instrument are considered highly relevant in testing psychometric properties. Cronbach’s alpha has been used as an indicator of reliability for decades, although alpha is known as being a lower bound to reliability and has been subject to critique (Revelle & Zinbarg, 2009; Sijtsma, 2009). Further, one important asset of an instrument, construct validity, has not been examined adequately yet. A review on the psychometric aspects of the SDQ (Stone et al., 2010) showed that mostly inadequate factor analytical methods and estimation methods were used in assessing the parent and teacher version. Therefore, the present study aims to assess reliability using a different indicator than alpha and to thoroughly analyze the factor structure of the parent version of the SDQ.

Since the 1950’s Cronbach’s alpha has been used as the solitary indicator for reliability, although many statisticians have pointed to the limited usefulness of alpha (Schmitt, 1996; Sijtsma, 2009). The use of alpha is described as problematic for two psychometric reasons. First, alpha is a lower bound to reliability and in many cases a gross underestimate. Second, alpha is often used as being an indicator of internal consistency leading researchers to conclusions that when alpha is acceptable, items must measure the same underlying construct. However, alpha merely implies the degree of interrelatedness of items, which has little to do with the internal structure of a test (Sijtsma, 2009). Recently, several alternatives to alpha such as Revelle’s beta, the greatest lower bound (GLB) and McDonald’s omega have been proposed and examined on their merits. Of these coefficients, omega seems to provide the most accurate estimate of reliability (Revelle & Zinbarg, 2009). Omega is defined as the percentage of a test that measures one construct and is decomposed into a general factor g, factors common to some but not all items f, specific factors unique to each item s and random error e (McDonald, 1999; Zinbarg et al., 2005). The present study will test reliability of the SDQ by using omega as an alternative to alpha.

With the common practice of analyzing the construct validity of the SDQ some problems are encountered. Many studies used Principal Component Analysis (PCA), (Becker et al., 2006; Du, Kou, & Coghill, 2008; Hawes & Dadds, 2004; Kashala, Elgen, Sommerfelt, & Tylleskar, 2005; Muris, Meesters, & Van den Berg, 2003; Smedje...
et al., 1999). PCA is a technique for constructing components of items without extracting the error part of an item (e.g., Bentler & Kano, 1990) in contrast to factor analysis (FA) that uses a factor model separating a true part and an error part. FA is congruent with classical test theory (Crocker & Algina, 1986). The latent five-factor structure of the SDQ defined by Goodman (1997) was based on theoretical foundations of psychopathology (DSM-IV; American Psychiatric Association, 1994). Hence, confirmatory factor analysis (CFA), instead of exploratory techniques such as PCA or FA, is more suited. Finally, responses to the SDQ are often skewed with an overrepresentation of responses indicating the absence of problem behaviors. As a consequence, CFA’s with normal theory estimators like Maximum Likelihood (ML) are unsuited (Finney & DiStefano, 2006). The metric of the response scales of the items is more categorical ordered than interval level, and thus Asymptotically Distribution-Free (ADF) or Weighted Least Squares (WLS) estimators are more suited. However, both estimators require extremely large samples. Muthén (1984) therefore developed robust WLS-estimators requiring smaller sample sizes.

While support for the five-factor structure using CFA is growing substantially (Becker, Woerner, Hasselhorn, Banaschewski, & Rothenberger, 2004; Palmieri & Smith, 2007; Sanne et al., 2009; Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006; Van Roy, Veenstra, & Cienach-Aas, 2008), important questions regarding construct validity remain unanswered. Comparing groups is often the focus in developmental psychopathology, underscoring the need of testing measurement invariance. Measurement invariance implies that evidence for construct validity is equal across groups (e.g., Meredith, 1993; Vandenberg & Lance, 2000); and meaningful comparisons can be made. Configural invariance is supported if each common factor has an identical set of items across groups. This is the case if parents use the same items for example, for boys and girls to classify a construct implicating a similar pattern of zero and non-zero loadings across gender. Metric invariance is supported if factor loadings are identical across groups. This means that the corresponding factors have the same meaning in the different groups. Scalar invariance implies that item intercepts are invariant across groups. In that case, cross-gender differences in the means of the items are due to differences in the means of the underlying constructs. Generally, there is consensus on CFA being the most powerful and versatile approach to test for measurement invariance (Steenkamp & Baumgartner, 1998).

In the present study, methodological aspects and demographic variables are examined with regard to measurement invariance. First, alternatives to mail surveys are well established (i.e. telephone interviews) (Erhart, Wetzel, Krügel, & Ravens-Sieberer, 2009) or emerging quickly at the time (i.e. web-based surveys) (Rhodes, Bowie, & Hergenrather, 2003). Second, gender differences are well known in the field of child psychopathology (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; Ford, Goodman, & Meltzer, 2003), and separate norms are available for most assessment tools (Achenbach & Rescorla, 2001; Goodman, 2001). Third, several demographic variables have been associated with adverse child development, such as maternal age (e.g. Croen, Najjar, Fireman, & Grether, 2007), maternal educational level (e.g. Meltzer, Gatward, Goodman, & Ford, 2003) and sibship size (Jensen, Bloedau, Degroot, Ussery, & Davis, 1990). Therefore, measurement invariance is tested for these variables.

The present study examines reliability and construct validity of the SDQ. Invariance of the parent version of the SDQ is examined for survey method, gender, maternal age, maternal educational level and number of siblings. We focus on the parent version since this version is most widely used for research purposes (Stone et al., 2010). Reliability indicators are hypothesized to be higher than the usually reported values because we do not use the lower bound alpha. We expect to confirm the five-factor structure by testing with CFA. Measurement invariance is thought being supported for gender, maternal age, maternal education, sibship size, and survey method.

Materials and Methods

Participants
Participants were 1,484 mothers (M age = 41.53, SD = 4.11) of 9-12 year-olds (Hiemstra, Ringlever, Otten, Jackson, Van Schaijck, & Engels, 2009). Mothers were recruited via primary schools, media, and health professionals. Mothers were administered the SDQ for one of their children. The target children were 713 boys and 771 girls (M age = 10.11, SD = 7.8). The majority of mothers (98.2 %) were of Dutch origin. Regarding educational level, 48.6 % of the mothers had a preparatory college or university degree (high), 34.5 % an intermediate general education (medium) and 16.9 % a lower education (low). Children had 0-1 siblings in 59.9 % and 2 or more siblings in 40.1 % of the families.

Procedures
Mothers completed the SDQ as part of a questionnaire booklet (38.7 %) or a telephone interview (61.3 %). The questionnaire booklet was sent via post and returned by an enclosed envelope. Students from the Radboud University Nijmegen administered the telephone interview, which they were trained for.

Measures
The Dutch parent version of the SDQ was used (Widenfelt, Goedhart, Treffers, & Goodman, 2003). The questionnaire consists of five subscales; each containing five
items measuring emotional symptoms, conduct problems, hyperactivity-inattention, peer problems and prosocial behavior. Mothers rated their child on a 3-point scale ranging from 0 (not true) to 2 (certainly true). The scoring procedures are available online at www.sdqinfo.org.

Mothers could choose between eight optional educational levels; elementary school (1), lower education (2), lower professional education (3), secondary school (4), intermediate professional education (5), advanced secondary school (6), preparatory college (7) or university (8). Subsequently, educational level was grouped into low (1-4), medium (5) and high (6-8) categories. Maternal age was categorized as low, medium or high based on the frequency distribution of maternal age with each group representing 33.3 % of the sample. The number of siblings was asked for and then grouped into categories 0-1 and 2 or more siblings. Because few families fulfilled the criterion of 0 siblings, 0 and 1 sibling(s) were grouped together.

**Strategy for analysis**

McDonald (1978, 1999) proposed measures for reliability based on a factor model with one factor (here denoted as \( \omega_h \), \( h \) relates to the hierarchical or general factor; this measure is identical with the measure of Jöreskog) and based on a factor model with several factors (denoted as \( \omega_h \), \( t \) relates to the total number of factors) (McDonald, 1978, 1999), see also Revelle and Zinbarg (2009). However, we are not interested in the reliability of a set of factors together but in the reliability of one construct or factor. For this reason our interest is in \( \omega_h \), especially because this reliability measure can easily be computed by hand from standard SEM-packages. Moreover, because the factor model will be analyzed based on polychorical correlations and robust WLS-estimators, \( \omega_h \) will give a better estimate of the reliability than CFA based on covariances and normal theory based estimators. The reliability index \( \omega_h \) of a factor with \( k \) indicators (items) is defined as follows:

\[
\omega_h = \frac{(\sum_{j=1}^{k} \lambda_j)^2}{(\sum_{j=1}^{k} \lambda_j)^2 + \sum_{j=1}^{k} \delta_j^2}
\]

with \( \lambda_j \) equals the standardized factor loading of indicator \( j \) of a latent variable and \( \delta_j^2 \) the standardized unique variance \((= 1 - \lambda_j^2)\). For each factor a 1-factor solution must be obtained to calculate \( \omega_h \), see also Revelle and Zinbarg (2009). Then the factor loadings are summed and this sum must be squared \((= A)\). If the standardized unique variances are given in the output, these variances must be summed \((= B)\). If the unique variances are not given, they can be calculated for each indicator by hand as \( 1 \) - the squared loading. Dividing \( A \) by \( A + B \) gives \( \omega_h \). In Zinbarg, Revelle, Yovel and Li (2005) several reliability measures are compared. The measures \( \alpha \) and \( \omega_h \) are both lower bound estimates of the true reliability, but \( \alpha \) is always lower than \( \omega_h \) and \( \omega_h \) is lower or equals the true reliability (only in case of essential tau equivalence -equal factor loadings for all indicators- the two measures give identical results).

To evaluate the five-factor structure of the SDQ, we used MPLUS version 5 (Muthén & Muthén, 1998-2007). The Weighted Least Square parameter estimator with standard errors and Mean-adjusted chi-square test statistic (WLSM)-estimator were used. The distribution of each of the response scales of the items \( y \) is replaced by a continuous distribution having a probability curve derived from the normal distribution \( N(\mu, \sigma^2) \). Each category represents a percentage of the sample. The categories of the original distribution are replaced by thresholds in the normal distribution. Three \( (C) \) response categories are replaced by two \( (C - 1) \) thresholds. If for example the first (zero) category contains 90% of all cases, the threshold value will be .84 (being a z-value in the standard normal distribution, representing 90% of all cases until \( z = .84 \)). If the second (1) category contains 15% of all cases, the threshold value will become 1.645 (representing 15% of all cases between \( z = .84 \) and \( z = 1.645 \)). The remaining cases are automatically located above \( z = 1.645 \) (5%). In this way the three response categories of the original scales were replaced by two thresholds in the normal distribution. For each pair of items, a correlation can be computed based on these threshold values and are called polychorical correlations. Thresholds and polychorical correlations are the input for the CFA (Finney & DiStefano, 2006).

Model fit was assessed with various fit indices, including robust chi-square with estimated degrees of freedom (df), comparative fit index (CFI; Bentler, 1990) and root mean squared error of approximation (RMSEA; Byrne, 1998). Conventional goodness of fit criteria in CFA, however, may be too restrictive (Marsh, Hau, & Wen, 2004). The reason for this is that in CFA cross loadings are constrained to zero whereas in Exploratory Factor Analysis (EFA) cross loadings are allowed and thus also estimated. According to Marsh et al. (2004) it is almost impossible to get an acceptable fit (e.g. CFI > .90, RMSEA < .05) because criteria are very stringent. Measurement invariance was assessed first in terms of configural invariance. Configural invariance implies that a model with specified item clusters fits well in all groups, so for survey method, gender, maternal age, maternal educational level and number of siblings. Factor loadings and thresholds are free to vary in the configural invariance model. As stated above, the response scale of the SDQ is ordered categorical. Therefore, testing of metric and scalar invariance must be conducted by constraining factor loadings and thresholds simultaneously because item probabilities of the factor indicators are influenced by both types of parameters (Muthén & Muthén, 1998-2006, 399-400; Kim & Yoon, 2011).

Assuming that configural invariance is supported, metric and scalar invariance were evaluated simultaneously. We compared the configural invariance model with the model where factor loadings and thresholds are constrained to be equal across
groups. To test whether the constraints significantly deteriorated the model, a chi-square difference test was used according to the recommendations on the website of Mplus, see http://www.statmodel.com/chidiff.html. We used the program TRd.exe to perform this difference test (http://psyphz.psych.wisc.edu/~shackman/Enzmann_Software.html).

When the chi-square difference test is applied to nested models, strengths and weaknesses appear which are essentially comparable to when the chi-square test is applied to any single model, namely, the test is directly affected by sample size, and for larger samples trivial differences may become significant (Schermelleh-Engel, Moosbrugger, & Müller, 2003). For large sample sizes, the chi-square statistic provides a highly sensitive statistical test, but not a practical test, of model fit (Cheung & Rensvold, 2002). Therefore, statistical significance was examined, but in addition we also assessed the increment of the fit indices CFI and RMSEA between the constrained and the unconstrained model (ΔCFI and ΔRMSEA). In their simulation studies Cheung and Rensvold (2002) found that ΔCFI had desirable characteristics as test of invariance. This was not true for RMSEA. We follow their rule of thumb that a decrease in CFI of .01 or less indicates that invariance should not be rejected.

Results

Internal reliability: Alpha and Omega

Cronbach’s alpha coefficients were adequate for Hyperactivity-Inattention (.82), acceptable for Emotional Symptoms (.71) and Prosocial Behavior (.68), but low for Conduct Problems (.54) and Peer Problems (.58). In contrast, the omega coefficients which were calculated from the CFA results were .91, .82, .85, .74 and .79 respectively indicating that the five scales have acceptable to very good reliabilities.

Construct validity: Configural invariance

The five-factor model was tested allowing factor loadings and thresholds free to vary (Model 1). Table 1 shows that configural invariance was supported for child’s gender, survey method, maternal educational level, maternal age and number of siblings with CFI > .900 and RMSEA ≤ .074. Hence, across these variables the five factors thus consisted of the same item clusters.

Construct validity: Metric and Scalar invariance

Factor loadings and thresholds were constrained to be equal in order to test metric and scalar invariance (Model 2). Table 1 shows the fit indices. The difference χ²-test between Model 1 and 2 was significant for child’s gender (Δχ² (70) = 245.95, p = 0.000), survey method (Δχ² (70) = 225.58, p = 0.000), maternal age (Δχ² (140) = 206.48, p = 0.000), maternal educational level (Δχ² (140) = 251.76, p = 0.000) and number of siblings (Δχ² (70) = 102.58, p = 0.007). However, the CFI-values for these models did not decrease with the critical value of .01. A decrease of .01 would indicate that the model is non-invariant at the metric and scalar level. Our results thus indicate that measurement invariance is still supported (Cheung & Rensvold, 2002).

Table 1 Goodness-of-Fit Indices

<table>
<thead>
<tr>
<th>Model loadings and thresholds</th>
<th>Variable</th>
<th>χ²</th>
<th>df</th>
<th>p</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Free to vary</td>
<td>Gender</td>
<td>4267.56</td>
<td>530</td>
<td>.000</td>
<td>.941</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
<td></td>
<td>4291.95</td>
<td>600</td>
<td>.000</td>
<td>.942</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
<td>Survey method</td>
<td>2139.20</td>
<td>530</td>
<td>.000</td>
<td>.954</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
<td></td>
<td>2361.09</td>
<td>600</td>
<td>.000</td>
<td>.950</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
<td>Maternal age</td>
<td>2389.50</td>
<td>795</td>
<td>.000</td>
<td>.953</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
<td></td>
<td>2468.11</td>
<td>935</td>
<td>.000</td>
<td>.955</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
<td>Educational level</td>
<td>2869.41</td>
<td>795</td>
<td>.000</td>
<td>.942</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
<td></td>
<td>2973.33</td>
<td>935</td>
<td>.000</td>
<td>.943</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
<td>Sibship size</td>
<td>2031.99</td>
<td>530</td>
<td>.000</td>
<td>.953</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
<td></td>
<td>2052.67</td>
<td>600</td>
<td>.000</td>
<td>.957</td>
</tr>
</tbody>
</table>

Because support was found for metric and scalar invariance, a final CFA was conducted over all respondents (CFI = .953; RMSEA = .062). Table 2 presents factor loadings, which are adequately high, mostly above .6, though three factor loadings were around .5. Factor correlations are shown in Table 3. Conduct problems is highly related to hyperactivity-inattention and moderately related to peer problems and prosocial behavior. Peer problems are highly related to emotional problems, and moderately to prosocial behavior.
Discussion

The present study aimed to evaluate a relatively new indicator for reliability, to examine the five-factor structure, and to test for measurement invariance for the parent version of the SDQ. As hypothesized, reliability indicators across all subscales were substantially higher when using omega. As this study is the first to test omega for the SDQ, thereby showing that all scales reach acceptable reliability, this study is a unique contribution to the literature. Furthermore, tests for measurement invariance are infrequent, especially in the case of the SDQ. Therefore, this study as a whole adds to the psychometric literature on the SDQ.

Statisticians have reported about the limited usefulness of alpha, which may explain the differences in alpha versus omega (Green & Yang, 2009). First, an assumption required for the derivation of alpha, essential tau equivalence, is frequently violated in practice which leads to a negatively biased estimate of reliability. Essential tau equivalence implies that items should refer to one single factor, meaning that a test is unidimensional. In practice, tests are often multidimensional. The second assumption, which should hold for the derivation of alpha, is that of uncorrelated errors. Often, errors are not purely random and thus may be correlated to some degree. Violation of uncorrelated errors leads to liberal estimates of reliability, which can be biased substantively. As assumptions for deriving alpha are thus often not met (Sijtsma, 2009), researchers do themselves injustice in those cases by keeping using alpha. Fortunately, multiple articles by statisticians have been published which point to several alternatives to alpha (Revelle & Zinbarg, 2009; Sijtsma, 2009). Of these alternatives McDonald’s omega shows the best results for multidimensional tests and calculation from factor analysis is easy (Revelle & Zinbarg, 2009).

In addition, alpha only is a measure for interrelatedness, not for the internal structure of a test. Different tests of varying factorial composition may have the same

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Factor Loadings of the SDQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Symptoms</td>
<td>Factor loadings</td>
</tr>
<tr>
<td>Somatic</td>
<td>.49</td>
</tr>
<tr>
<td>Worries</td>
<td>.70</td>
</tr>
<tr>
<td>Unhappy</td>
<td>.79</td>
</tr>
<tr>
<td>Clingy</td>
<td>.80</td>
</tr>
<tr>
<td>Fears</td>
<td>.64</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td></td>
</tr>
<tr>
<td>Tempers</td>
<td>.61</td>
</tr>
<tr>
<td>Obedient*</td>
<td>.70</td>
</tr>
<tr>
<td>Fights</td>
<td>.66</td>
</tr>
<tr>
<td>Lies</td>
<td>.59</td>
</tr>
<tr>
<td>Steals</td>
<td>.45</td>
</tr>
<tr>
<td>Hyperactivity-Inattention</td>
<td></td>
</tr>
<tr>
<td>Restless</td>
<td>.83</td>
</tr>
<tr>
<td>Fidgety</td>
<td>.80</td>
</tr>
<tr>
<td>Distractible</td>
<td>.91</td>
</tr>
<tr>
<td>Reflective*</td>
<td>.60</td>
</tr>
<tr>
<td>Persistent*</td>
<td>.93</td>
</tr>
<tr>
<td>Peer Problems</td>
<td></td>
</tr>
<tr>
<td>Solitary</td>
<td>.50</td>
</tr>
<tr>
<td>Good friend*</td>
<td>.63</td>
</tr>
<tr>
<td>Popular*</td>
<td>.82</td>
</tr>
<tr>
<td>Bullied</td>
<td>.70</td>
</tr>
<tr>
<td>Best with adults</td>
<td>.62</td>
</tr>
<tr>
<td>Prosocial Behavior</td>
<td></td>
</tr>
<tr>
<td>Considerate</td>
<td>.86</td>
</tr>
<tr>
<td>Shares</td>
<td>.68</td>
</tr>
<tr>
<td>Caring</td>
<td>.77</td>
</tr>
<tr>
<td>Kind to kids</td>
<td>.59</td>
</tr>
<tr>
<td>Helps out</td>
<td>.71</td>
</tr>
</tbody>
</table>

Table 3: Correlations between the Model Factors

<table>
<thead>
<tr>
<th>Table 3 Correlations between the Model Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1 Emotional Symptoms</td>
</tr>
<tr>
<td>2 Conduct problems</td>
</tr>
<tr>
<td>3 Hyperactivity-Inattention</td>
</tr>
<tr>
<td>4 Peer problems</td>
</tr>
<tr>
<td>5 Prosocial Behavior</td>
</tr>
</tbody>
</table>

Note. All correlations are significant at the $p < .001$ level.
alpha values, implying that alpha says very little about multiple-factor item structures (Gjitsma, 2009). If scholars wish to examine the internal structure of a test, factor analysis is very suitable.

In accordance with previous research, the proposed five-factor structure was confirmed (Becker et al., 2004; Palmieri & Smith, 2007; Sanne et al., 2009; Van Leeuwen et al., 2006; Van Roy et al., 2008). The SDQ showed to be invariant on a configurational, scalar and metric level for survey method, child’s gender, mental educational level, mental age and number of siblings. In the preponderance of literature on the SDQ, three studies examined measurement invariance of the SDQ parent version, wherein evidence for invariance was found for age, gender, ethnic groups, maternal educational level and different informants (Hill & Hughes, 2007; Palmieri & Smith, 2007; Sanne et al., 2009). Measurement invariance is now thus established for several demographic factors, which implies that when studying these factors in relation to the SDQ, group comparisons can be meaningful.

A few limitations should be noted. First, only the parent version of the SDQ was used. It would be very interesting to examine omega for other versions of the SDQ. Future research should reveal whether omega is consistently a better measure for reliability than alpha, and could possibly serve as a successor to alpha. Regarding invariance, except for one study that tested measurement invariance of the teacher version of the SDQ (Zwirs et al., 2010), to our knowledge no research into invariance on the teacher or self-report version has been conducted. Further, reliability and factor analysis are just two aspects of the psychometric properties of an instrument. Other aspects of validation, such as the predictive validity of the SDQ, also warrant investigation (Kane, 2006).

The present study has an important implication for research and practice. As the SDQ is often used as a screening instrument (e.g., Glazebrook, Hollis, Heussler, Goodman, & Coates, 2002), it is essential that norms are based on meaningful group differences between, for example, boys and girls. At present, norms have only been constructed for six countries (www.sdqinfo.org). So, when constructing norms for other countries where the SDQ is used frequently, researchers are encouraged to test for measurement invariance. Especially when dealing with child psychopathology, researchers should not risk their conclusions based on the SDQ to be at best ambiguous or at worst erroneous, because they neglected measurement invariance.

References


Chapter 4

The Strengths and Difficulties Questionnaire: Psychometric Properties of the Parent and Teacher Version in Children Aged 4-7

Submitted for publication:
CHAPTER 4

PSYCHOMETRICS SDQ PARENT AND TEACHER VERSION

Abstract

The Strengths and Difficulties Questionnaire is one of the most employed screening instruments. Although there is a large research body investigating its psychometric properties, reliability and validity are not yet fully tested using modern techniques. Therefore, we investigate reliability, construct validity, measurement invariance, and predictive validity of the parent and teacher version in children aged 4-7. Besides, we intend to replicate previous studies by investigating test-retest reliability and criterion validity. In a Dutch community sample 2,238 teachers and 1,513 parents filled out questionnaires regarding problem behaviors and parenting, while 1,831 children reported on sociometric measures at T1. These children were followed-up during three consecutive years. Reliability was examined using Cronbach’s alpha and McDonald’s omega, construct validity was examined by Confirmatory Factor Analysis, and predictive validity was examined by calculating developmental profiles and linking these to measures of inadequate parenting, parenting stress and social preference. Further, mean scores and percentiles were examined in order to establish norms. Omega was consistently higher than alpha regarding reliability. Omega indices were all adequate, while alpha was not for some subscales. The original five-factor structure was replicated, and measurement invariance was established on a configural level. Further, higher SDQ scores were associated with future indices of higher inadequate parenting, higher parenting stress and lower social preference. Finally, previous results on test-retest reliability and criterion validity were replicated. This study is the first to show SDQ scores are predictively valid, attesting to the feasibility of the SDQ as a screening instrument. Future research into predictive validity of the SDQ is warranted.

Introduction

In child mental health care and research, screening instruments play an important role in measuring what types of psychosocial problems and strengths may be identified and how severe these problems are, if any. The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is one of the most widely used screening instruments for these purposes. Although much research has been conducted into reliability and validity of the SDQ, several issues warrant further investigation. First, although reliability has been extensively studied (see for a review Stone, Otten, Engels, Vermulst, & Janssens, 2010), reliability of the subscales seems insufficient, specifically for the conduct problems and peer problems scales. Second, construct validity and measurement invariance have not been examined frequently for both the parent and teacher version, nor for younger children. Third, while stability of SDQ scores over time has been reported (Hawes & Dadds, 2004; Perren, Stadelmann, Von Wyl, & Von Klitzing, 2007), the degree to which SDQ scores predict subsequent maladjustment has not been examined previously. The goal of the present study was to investigate these three issues. In addition, we present Dutch normative data for the parent and teacher version of the SDQ and report on test-retest reliability and criterion validity.

Regarding reliability, mostly Cronbach’s alphas have been reported (see Stone et al., 2010). Recently, the use of this reliability coefficient has been subject to critique according to psychometricians, due to its underestimation of reliability (Revelle & Zinbarg, 2009; Sijtsma, 2009), specifically when response scales of items have few categories and when scale distributions are skewed (Muthén, 1984). Evidently, this occurs frequently if not always when measuring psychopathology. Therefore, alternatives to alpha have been suggested and tested, with McDonald’s omega or Jöreskog rho being the most accurate (Revelle & Zinbarg, 2009). Employing these accurate measures seems imperative when testing reliability (cf. Schweizer, 2011). Indeed, it has been found that omega coefficients yield higher estimates for the SDQ than alpha (Gómez-Beneyto et al., 2013; Ezpeleta, Granero, de la Osa, Penelo, & Domínech, 2013; Kóbor, Takács, & Urbán, 2013; Stone et al., 2013). Still, these studies are limited by investigating solely the parent version (Gómez-Beneyto et al., 2013; Stone et al., 2013), relatively small sample sizes (Kóbor et al., 2013), and a limited age range, namely preschoolers (Ezpeleta et al., 2013).

Second, support for the SDQ’s five-factor structure is growing as studies increasingly employ confirmatory factor analysis to test its hypothesized factor structure. This is the case for both the parent and teacher version, and for various age ranges (Becker, Woerner, Hasselhorn, Banaschewski, & Rothenberger, 2004; Moriwaki & Kamiya, 2014; Palmieri & Smith, 2007; Sanne, Torsheim, Heiervang, & Stormark, 2009; Van Roy, Veenstra, & Clerch-Aas, 2008), with only two studies
examining this in children aged 4-7 specifically (Niclasen, Skovgaard, Andersen, Sormhovd, & Obel, 2013; Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006). Also, relatively few studies have tested for measurement invariance, namely whether the underlying structure is identical across different groups. Three studies tested measurement invariance for the parent version in older age groups (Palmiéri & Smith, 2007; Sanne et al., 2009; Stone et al., 2013), and two studies in children aged 4-7 (Hill & Hughes, 2007; Niclasen et al., 2013). These studies found the SDQ to be invariant across gender, age, ethnicity, and maternal education. Regarding the teacher version, two studies tested for measurement invariance in older age groups (Ruchkin, Koposov, Vermeiren, & Schwab-Stone, 2012; Sanne et al., 2009), and three studies in children aged 4-7 (Hill & Hughes, 2007; Niclasen et al., 2013; Zwirs et al., 2011). These studies found the SDQ to be invariant across ethnicity, but results are inconsistent regarding gender. Due to the limited number of studies reporting on construct validity and measurement invariance for children aged 4-7 and the inconsistent results on measurement invariance for the teacher version, it was deemed important to investigate these issues in the present study. Measurement invariance is investigated for gender, age, and ethnicity.

Finally, to our knowledge predictive validity has not been investigated for the SDQ. It has been found that SDQ scores predict SDQ scores over a one-year interval (Hawes & Dadds, 2004; Perren et al., 2007), for both the parent and teacher version. Still, these results do not evidence that SDQ scores are related to a criterion measure over time, they merely show that SDQ scores are correlated over time. Therefore, it was deemed important to investigate the SDQ’s predictive validity in relation to two factors related to child psychopathology; maladaptive parenting and social preference. Specifically, we hypothesized that higher SDQ scores would predict maladaptive parenting and higher parenting stress for the parent version and that higher SDQ scores would predict lower levels of social preference (i.e., the degree to which a child is liked by classmates) for the teacher version.

In the Netherlands, the SDQ is increasingly used to assess psychosocial problems in children, however normative data on Dutch SDQ scores are limited by a small sample size and selectiveness of the sample (Goedhart, Treffers, & Van Widenfelt, 2003; Van Vuuren, Diepenmaat, Reijneveld, & Van der Wal, 2008; Vogels, Crone, Hoekstra, & Reijneveld, 2005). Therefore, in this paper Dutch normative data are presented for both the parent and teacher version and based on a relatively large sample. In addition, we examined criterion validity in order to replicate previous studies, by comparing SDQ scores to scores obtained by the Child Behavior Check List and Teacher Report Form scores (Achenbach & Rescorla, 2000). Similarly, we examined criterion validity for replication purposes for the parent version. Regarding the teacher version, criterion validity has not been extensively investigated (Stone et al., 2010). Therefore, we sought to validate the SDQ teacher version by using measures proximal to teachers. Sociometric measures may be particularly useful in this respect, as these may reflect difficulties in peer relations, behaviors exhibited within the school context and are related to child psychopathology (e.g., Van Lieër & Koot, 2010).

In sum, the present study examined reliability (i.e., Cronbach’s alpha, McDonald’s Omega), test-retest reliability, as well as construct, criterion (concurrent and predictive) validity and measurement invariance of both the parent and teacher version of the SDQ for children aged 4-7. We expected that omega values would yield higher reliability coefficients than alpha. Next, we expected to confirm the hypothesized five-factor structure, to find invariance for gender, age and ethnicity, and we expected substantial inter-correlations among SDQ subscales. Further, we expected that SDQ scores inter-correlate over a retest interval, correlate with similar measures of psychopathology, and are related to maladaptive parenting and sociometric measures within and over time. Finally, we present Dutch normative data for children aged 4-7.

Method

Participants and Procedure

In the 2008-2009 school year, schools were randomly selected from all elementary schools in the Netherlands. Schools in the larger counties (i.e., Noord-Holland, Zuid-Holland, Noord-Brabant, and Gelderland), as well as in the four largest cities (i.e., Amsterdam, Rotterdam, Den Haag, and Utrecht), were oversampled. A total of 440 schools were selected. Directors received a letter in which they were invited to participate in the study. Subsequently, they were called to ask whether they wanted to participate. Directors of 29 schools (6.6%) promised their cooperation. These 29 schools together account for approximately 2300 pupils from the groups 1 to 4. At the initial measurement, during the 2009-2010 school year, teachers completed the SDQ concerning 2,238 pupils. Regarding the second and third measurement, SDQ data were collected through the teachers about 1,962 and 1,572 pupils, respectively. At the three annual measurement occasions, SDQ data were also collected by means of the parents of the pupils, concerning 1,513, 1,036, and 888 children. Again, at all three annual measurement occasions, sociometric interviews were held with the children themselves, concerning 1,831, 1,603, and 1,770 children. From all these children, 25% came from each of the four groups, and half of the cases concerned boys. Of all children, 79.5% had parents who were both born in the Netherlands, whereas 20.5% had at least one parent who was born abroad (3.5% of Turkish origin, 5.4% Moroccan, and 1.9% Surinam; the remaining children came from parents born in a wide variety of countries). Finally, parents and teachers filled out another SDQ 6 weeks after T1 for 203 and 188 randomly chosen children, respectively, in order to examine test-retest reliability.
Measures

Strengths and Difficulties Questionnaire. The Dutch parent and teacher informant version of the SDQ was used at all waves (SDQ; Van Widenfelt, Goedhart, Treffers, & Goodman, 2003). The questionnaire consists of five subscales, each of which contain five items measuring emotional symptoms (e.g., many fears, easily scared), conduct problems (e.g., often lies or cheats), hyperactivity-inattention (e.g., restless, overactive, cannot stay still for long), peer problems (e.g., picked on or bullied by other children), and prosocial behavior (e.g., considerate of other people’s feelings). Parents and teachers rated children on a 3-point scale ranging from 0 (not true) to 2 (certainly true). The scoring procedures are available online at http://www.sdqinfo.org.

For each of the five subscales, a score ranges from 0-10 if all five items were completed. Further, a total difficulties score can be calculated by summing the scores from the first four subscales (range 0-40). Mean scores on the SDQ parent version at all measurements in this sample are relatively low for the emotional symptoms scale (range $M = 1.60, SD = 1.81 - M = 1.67, SD = 1.87$), conduct problems scale (range $M = 1.02, SD = 1.37 - M = 1.28, SD = 1.44$), hyperactivity scale (range $M = 2.96, SD = 2.57 - M = 2.98, SD = 2.50$), peer problems scale (range $M = 9.8, SD = 1.39 - M = 1.08, SD = 1.43$), and total difficulties scale (range $M = 6.68, SD = 5.26 - M = 6.93, SD = 4.85$), and relatively high for the prosocial scale (range $M = 8.16, SD = 1.72 - M = 8.52, SD = 1.66$). This also holds for the teacher version; emotional symptoms scale (range $M = 1.03, SD = 1.59 - M = 1.44, SD = 1.89$), conduct problems scale (range $M = 7.4, SD = 1.42 - M = 8.2, SD = 1.31$), hyperactivity scale (range $M = 2.64, SD = 2.83 - M = 2.89, SD = 2.95$), peer problems scale (range $M = 1.05, SD = 1.51 - M = 1.22, SD = 1.65$), and total difficulties scale (range $M = 5.58, SD = 4.86 - M = 6.27, SD = 5.63$), and relatively high for the prosocial scale (range $M = 7.67, SD = 2.35 - M = 8.10, SD = 2.13$). In conclusion, psychosocial difficulties in children between the ages of 4 and 7 are limited in this sample. In fact, we could extend this conclusion to 8 and 9 year-olds, since the oldest children had reached that age by the third measurement.

Child Behavior Check List (CBCL/1-5) and (Caregiver-)Teacher Report Form. The Dutch versions of the CBCL/1-5, CBCL, C-TRF and TRF were used to assess internalizing and externalizing behavior as reported by parents and teachers at T1 (Achenbach & Rescorla, 2000; Achenbach & Rescorla, 2001; Verhulst, Koot, Akkerhuis, & Veerman, 1990; Verhulst, Van der Ende, & Koot, 1997). The CBCL/1-5/C-TRF, used for children aged 1.5-5 years, comprises 100 items; the CBCL/TRF targets 5-18-year-olds and consists of 118 items. These items are rated using a 3-point Likert scale, where 0 indicates responses of “not true”, 1 “somewhat or sometimes true”, and 2 “very true or often true”. In all four versions, scores can be calculated regarding internalizing, externalizing and total behavioral problems (Verhulst et al., 1997). The distributions of the scores were skewed, and therefore scores above the 99th percentile were rescaled to the 99th percentile value. Cronbach's alphas ranged from .84-.87 for the internalizing scale, from .87-.93 for the externalizing scale, and from .91-.94 for the total problems scale, for the parent and teacher version for younger and older children.

Parenting Daily Hassles. At all waves parents rated the frequency of daily hassles with their child over the past 6 months (PDH; Crnic & Booth, 1991; Van der Wal, Van Eijsden, & Bonsel, 2007). The questionnaire consists of 20 events of which the parent has to rate how often they occur (seldom, sometimes, often, constantly). A mean score was calculated with higher scores indicating higher parenting stress. Psychometric properties of the PDH have been found adequate (Crnic & Booth, 1991). Cronbach’s alphas were .77, .79, and .78 at T1, T2, and T3.

The Parenting Scale. The Parenting Scale was used at all waves and asks parents to rate 30 short parenting situations on a 7-point scale (TPS; Arnold, O’Leary, Wolff, & Acker, 1993). Sample items include “When I want my child to stop doing something I firmly tell my child to stop / I coax or beg my child to stop” and “When I’m upset or under stress I am picky and on my child’s back / I am no more picky than usual”. Inadequate parenting behavior is divided across three subscales: permissiveness, restrictiveness, and verbosity. All the items sum up to the total score, which was used in the current study. Higher scores reflect more inadequate parenting behavior. Psychometric properties are adequate (Arnold et al., 1993). Cronbach’s alphas were .77, .81, and .80, for the total score at T1, T2 and T3.

Social Preference. At all waves children were interviewed individually. During these interviews, children were shown a photograph of their classmates. A trained research assistant pointed out a child on the photograph and asked the child whether (s)he knew who this child was, ensuring familiarity, and was then asked whether (s)he liked, disliked the child or thought neutral of him/her. To increase comprehension and ease shy children, the child could respond verbally or by pointing to three fluffy smileys, with either a happy, sad or neutral expression. This procedure was repeated until the child gave a nomination about every child in the class. The order of asking questions about children in the photograph was counterbalanced, such that the interviewer started either at the upper left, upper right, lower left or lower right corner of the photograph. Unlimited nominations (like, dislike, neutral) were used, because these tend to spread more evenly among children in a class than limited nominations (i.e., fewer children receive a raw nomination score of zero). For each child, scores were calculated that indicate the extent to which a child is liked by fellow pupils (“Like-score”), and the extent to which fellow pupils do not like the child (“Dislike-score”). These scores were standardized within each classroom. The total least-liked nomination was subtracted from the total most-liked nomination to obtain a measure of social preference (cf. Coie, Dodge, & Coppotelli, 1982). These scores were obtained at T1, T2, and T3.
Strategy for Analysis

For the SDQ, we computed the reliability measure of Cronbach’s alpha. Also, we computed rho of Jöreskog (Jöreskog, 1971), also known as McDonald’s omega (McDonald, 1978, 1999). This measure shows the relationship between the variance explained by a factor and the total amount of variance to be explained by that factor, and has been recommended to be used (Schweizer, 2011). Research in which omega is applied to the SDQ, has shown good results (Kuijpers, Otten, Vermulst, & Engels, 2014; Stone et al., 2013). Reliability measures less than 0.70 are considered moderate, reliability measures between 0.70 and 0.80 are regarded sufficient, and measures above 0.80 are good (Evers, Lucassen, Meijer, & Sijtsma, 2010). Furthermore, we computed Spearman’s rho correlations between SDQ scales at T1 and SDQ scales completed after a retest interval of 6 weeks in order to examine test-retest reliability.

Construct validity was examined using confirmatory factor analysis (CFA). By means of CFA, it was tested whether the assumed five factor model of the SDQ could be confirmed, using Mplus (Muthén & Muthén, 1998-2007). For brevity reasons, for a detailed description of our analytical strategy regarding CFA we refer to Stone et al. (2013). Model fit was assessed with various fit indices, including robust chi-square with estimated degrees of freedom (df), comparative fit index (CFI; Bentler, 1990), and root mean squared error of approximation (RMSEA; Byrne, 1998). It is assumed that a factor model has a good fit when CFI > .95 en RMSEA < .05 and is acceptable when CFI > .90 en RMSEA < .08 (Marsh, Hau, & Wen, 2004).

Criterion validity is present when the score corresponding to an instrument is related to the score on an external criterion (an existing valid instrument) that measures the same property. The SDQ is valid when scores on the SDQ correlate sufficiently high with scores produced by other instruments that also measure psychosocial problems in children. Correlations < .30 are considered low, ≥ .30 average/medium, and ≥ .50 high (Cohen, 1992).

To investigate the predictive validity of the SDQ, we used Growth Mixture Modeling (GMM) (Muthén & Muthén, 1998-2007). By means of GMM, developmental profiles can be established, based on the SDQ scores at the three points in time. By doing so, we considered the development of the SDQ scores over time, instead of studying a single score at one moment in time. These profiles are constructed on the basis of growth parameters of the SDQ scores over the three measurements. In this case, these growth parameters consist of the intercept and the slope. The intercept can be regarded as the initial level of the SDQ scores. The slope represents the degree of change of these scores over time. To investigate the number of different profiles that are present in the population to be studied, we examined the most obvious ‘solution’, according to the fit statistics and theory. Several fit statistics are available, on the basis of which the best fitting number of profiles can be determined: The BIC (Bayesian Information Criterion), and the AIC (Akaike Information Criterion) (Burnham & Anderson, 2004). The model presenting the lowest value shows the best fit. The entropy value shows a good fit when being equal to or above 0.80. Subsequent to the identification of developmental profiles, one-way univariate ANOVA’s were conducted to test whether these groups differed on parenting measures and social preference scores.

Results

Reliability

The results with respect to reliability are presented in Table 1. Cronbach’s alpha ranges from .46 to .82 for the parent version, and from .53 to .88 for the teacher version. McDonald’s omega ranges from .67 to .90 for the parent version, and from .82 to .93 for the teacher version. We may conclude that the reliability indexed by Cronbach’s alpha is insufficient for the conduct problems, peer problems, emotional symptoms and prosocial scales of the SDQ parent version, while reliability indexed by McDonald’s omega yields sufficient to good estimates for all subscales. Reliability indexed by Cronbach’s alpha of the teacher version is insufficient for the conduct problems and peer problems scales, and good for all subscales when indexed by McDonald’s omega.

Furthermore, test-retest reliability of the parent version was examined, with correlations of .77 for the total problems scale, .81 for hyperactivity-inattention, .72 for emotional problems, .72 for prosocial behaviour, .54 for peer problems and .55 for conduct problems. For the teacher version, correlations of .80 for the hyperactivity-inattention and total problems scales, .77 for emotional problems, .70 for prosocial behaviour, .65 for peer problems and .58 for conduct problems were found. All correlations were significant at p < .001.

Construct Validity

It was examined whether the meaning of the five SDQ subscales is equivalent across several important characteristics (i.e., gender, age, and ethnicity), which is referred to as measurement invariance. It is not intended that the meaning of, for example, Emotional symptoms, is different for the 4-5 year olds than for the 6-7 year olds. The procedure applied and the corresponding outcomes are specified in Appendix 1. Based on the outcomes, we may conclude that the construct validity is not different regarding gender, age, and ethnicity, for the parent version of the SDQ. The comparison between boys and girls, older and younger children, and native and
The factor loadings are adequate, that is to say, larger than or equal to .40, although a few of the SDQ thus has an acceptable fit. This means that the five theoretically supposed scales are empirically demonstrable in this case as well. Again, robustness of the factor structure is demonstrated by showing that the structure is identified at three time-points. Standardized loadings are reported in Table 2.

Table 3 displays the mutual correlations between the latent five factors with regard to the parent and teacher version of the SDQ. As expected, the four SDQ problem scales are positively correlated, where the strongest correlation is found between Conduct problems, Hyperactivity and Peer problems. The problem scales are negatively correlated with Prosocial behavior, indicating that more problem behavior is associated with less prosocial behavior.

Criterion Validity: Correlations between SDQ and CBCL/TRF

The scores on the CBCL/TRF scales are correlated with the scores on the SDQ scales. Results of the parent and teacher version are presented in Table 4. From this table, one can deduce that the SDQ Total Difficulties scores correlate strongly with the CBCL and TRF total problems scores. The SDQ subscale Emotional symptoms correlates highly with the Internalizing problems scale as measured by the CBCL and TRF. The SDQ scales that point to externalizing problem behavior (Conduct problems, Peer problems, and Hyperactivity) are closely related to the CBCL and TRF Externalizing problems scale. All of these high correlations indicate a high degree of SDQ criterion validity.

Criterion Validity: Correlations among SDQ subscales and SDQ scales with parenting measures

First, we examined whether the subscales of the parent and teacher version are correlated. We found low but significant (p < .01) correlations for Emotional symptoms (.26), Conduct problems (.29), and Prosocial behavior (.21), and medium for Peer problems (.32), Hyperactivity (.48) and Total difficulties (.40).

Second, we examined whether SDQ scores were related to scores associated with psychosocial problems. It was expected that as parents raise their children more inadequate, these children would score higher on the SDQ problem scales. Obviously, this hypothesis especially concerned the parent version of the SDQ, yet we also checked whether high scores on inadequate parenting behavior were related to high stomach-aches, or nausea' (somatic) from the Emotional symptoms scale, and 'Steals from home, school or elsewhere' (steals) from the Conduct problems scale.

The fit of the CFA with regard to the teacher version was \(\chi^2(265) = 2619.55, p = 0.000, \text{CFI} = .920, \text{RMSEA} = .063\) at the first measurement, \(\chi^2(265) = 2930.75, p = 0.000, \text{CFI} = .930, \text{RMSEA} = .071\) at the second measurement, and \(\chi^2(265) = 2330.38, p = 0.000, \text{CFI} = .930, \text{RMSEA} = .070\) at the third measurement. Like the parent version, the teacher version of the SDQ has an acceptable fit. This means that the five theoretically supposed scales are empirically demonstrable in this case as well. Again, robustness of the factor structure is demonstrated by showing that the structure is identified at three time-points. Standardized loadings are reported in Table 2.
### Table 2  Factor loadings of the parent and teacher version of the SDQ

<table>
<thead>
<tr>
<th>Factor loadings</th>
<th>Parent</th>
<th>Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
</tr>
<tr>
<td><strong>Emotional symptoms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Somatic</td>
<td>.39</td>
<td>.46</td>
</tr>
<tr>
<td>Worry</td>
<td>.72</td>
<td>.77</td>
</tr>
<tr>
<td>Unhappy</td>
<td>.84</td>
<td>.78</td>
</tr>
<tr>
<td>Clingy</td>
<td>.66</td>
<td>.72</td>
</tr>
<tr>
<td>Fears</td>
<td>.63</td>
<td>.67</td>
</tr>
<tr>
<td><strong>Conduct problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tantrums</td>
<td>.59</td>
<td>.64</td>
</tr>
<tr>
<td>Obedient*</td>
<td>.61</td>
<td>.60</td>
</tr>
<tr>
<td>Fights</td>
<td>.73</td>
<td>.61</td>
</tr>
<tr>
<td>Lies</td>
<td>.52</td>
<td>.55</td>
</tr>
<tr>
<td>Steals</td>
<td>.35</td>
<td>.28</td>
</tr>
<tr>
<td><strong>Hyperactivity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restless</td>
<td>.77</td>
<td>.82</td>
</tr>
<tr>
<td>Fidgeting/squirming</td>
<td>.73</td>
<td>.73</td>
</tr>
<tr>
<td>Distracted</td>
<td>.80</td>
<td>.82</td>
</tr>
<tr>
<td>Thinks*</td>
<td>.62</td>
<td>.67</td>
</tr>
<tr>
<td>Completes*</td>
<td>.76</td>
<td>.79</td>
</tr>
<tr>
<td><strong>Peer problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solitary</td>
<td>.51</td>
<td>.45</td>
</tr>
<tr>
<td>Good friend*</td>
<td>.44</td>
<td>.57</td>
</tr>
<tr>
<td>Popular*</td>
<td>.81</td>
<td>.84</td>
</tr>
<tr>
<td>Bullied</td>
<td>.61</td>
<td>.57</td>
</tr>
<tr>
<td>Good with adults</td>
<td>.59</td>
<td>.63</td>
</tr>
<tr>
<td><strong>Prosocial behavior</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Considerate</td>
<td>.79</td>
<td>.89</td>
</tr>
<tr>
<td>Shares</td>
<td>.59</td>
<td>.68</td>
</tr>
<tr>
<td>Helpful</td>
<td>.59</td>
<td>.62</td>
</tr>
<tr>
<td>Kind</td>
<td>.52</td>
<td>.61</td>
</tr>
<tr>
<td>Helps</td>
<td>.57</td>
<td>.55</td>
</tr>
</tbody>
</table>

Note: Items marked with an asterisk are reversed items.

### Table 3  Correlations between the latent SDQ subscales of the parent and teacher version

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional symptoms</td>
<td>.41</td>
<td>.28</td>
<td>.54</td>
<td>.38</td>
<td>.42</td>
<td>-20</td>
<td>-23</td>
<td>-43</td>
<td>-37</td>
<td>-57</td>
</tr>
<tr>
<td>Conduct problems</td>
<td>.31</td>
<td>.69</td>
<td>.65</td>
<td>.76</td>
<td>.66</td>
<td>-27</td>
<td>-26</td>
<td>-78</td>
<td>-60</td>
<td>-82</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.57</td>
<td>.74</td>
<td>.62</td>
<td>.40</td>
<td>.54</td>
<td>-80</td>
<td>-78</td>
<td>-10</td>
<td>-84</td>
<td>-92</td>
</tr>
<tr>
<td>Peer problems</td>
<td>.64</td>
<td>.62</td>
<td>.42</td>
<td>.62</td>
<td>.44</td>
<td>-69</td>
<td>-81</td>
<td>-22</td>
<td>-33</td>
<td>-37</td>
</tr>
<tr>
<td>Prosocial behavior</td>
<td>-21</td>
<td>-64</td>
<td>-46</td>
<td>-36</td>
<td>-45</td>
<td>-33</td>
<td>-60</td>
<td>-22</td>
<td>-33</td>
<td>-37</td>
</tr>
</tbody>
</table>

Note. Correlations above the diagonal represent those for the teacher version, below for the parent version. All correlations are significant at $p < .001$. 
scores on the SDQ problem scales of the teacher version. If we would find these correlations, than that too would be indicative of the criterion validity of the SDQ teacher version. In Table 5, correlations between the SDQ scores and scores on the TPS and PDH are presented. All subscales of the SDQ parent version are significantly correlated with the TPS scores (range .13 - .24) and with the PDH scores (range .22 - .40). Highest correlations were found between Total difficulties and the TPS- and the PDH-score, respectively .24 and .40. It appears that the less adequate parents raise their children, the more problems these children exhibit, and that the more problems children exhibit, the greater parents’ daily hassles tend to be. The SDQ scales of the teacher version are hardly associated with the TPS scores. However, these scales are associated with PDH scores. The correlations are low, albeit in the expected direction. As children are experienced by their teachers as more problematic, parents of these children experience more daily hassles.

Finally, the SDQ’s criterion validity was examined by relating SDQ scores to like, dislike and social preference scores. These three scores correlate -.41, .42, and -.43, respectively, with the SDQ Total Difficulties score of the teacher version, and -0.29, 0.26, and -0.29 with the Total Difficulties score of the parent version. Equivalent correlations apply to the SDQ subscales (see Table 5). Hence, it appears that as pupils exhibit more psychosocial problems, they are less liked by their classmates. In conclusion, we may state that – with the findings above – the criterion validity of the SDQ is amply demonstrated.

Criterion Validity: Predictive Validity

Finally, the predictive validity was studied by examining whether developments in the course of SDQ scores over three measurements, were predictive for the course of inadequate parenting behavior and daily parenting hassles for the parent version, and were predictive of social preference scores for the teacher version, over the same period of time. Predictive validity is present when SDQ scores are predictive of scores on these parenting and social preference measures.

At the first step, we tested which model fitted the data best, using GMM. As can be seen in Table 6, when taking all fit statistics in consideration (i.e., relatively low levels of the AIC and BIC combined with a good entropy), these call for a model providing three developmental pathways. One large group scores consistently low on the SDQ total score (85.7%); one group scores high and demonstrates a slight decrease over time (5.1%); and one group that starts somewhat lower than the previous group, but shows a small increase over time (9.1%). These pathways are illustrated in Figure 1.

At the second step, the developmental pathways were linked to scores on TPS and PDH. Results are presented in Table 7. The findings show that developmental pathways of the SDQ are associated with scores on TPS, with significantly higher

### Table 4: Correlations between SDQ scores of the parent version and CBCL scores

<table>
<thead>
<tr>
<th>SDQ</th>
<th>CBCL 6-18 years old</th>
<th>TRF 6-18 years old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional symptoms</td>
<td>CBCL</td>
<td>TRF</td>
</tr>
<tr>
<td>Conduct problems</td>
<td>.73</td>
<td>.28</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>.37</td>
<td>.41</td>
</tr>
<tr>
<td>Peer problems</td>
<td>.40</td>
<td>.40</td>
</tr>
<tr>
<td>Total Difficulties</td>
<td>-.18</td>
<td>-.32</td>
</tr>
</tbody>
</table>

Note: All correlations are significant at p < .001.
scores in the high-decreasing group as compared to the stable-low group. At time three there was an overall significant effect ($p = .045$). However post-hoc tests (Bonferroni) revealed no significant differences between the different groups.

Regarding daily hassles, at the time of the first measurement, the three groups all differed significantly from each other. In the second and third measurement only the stable-low group and the high-decreasing group differed significantly. Strikingly, the two high trajectories hardly differ from each other with regard to the parenting measures, differences mainly exist between the large group exhibiting few problems and the two high trajectories. Less inadequate parenting behavior occurs and less daily hassles are experienced in the group exhibiting few problems, as compared to the other two groups. In sum, we can conclude that the SDQ demonstrates predictive validity in a sense that higher levels of psychopathology over time are generally associated with more parenting problems and daily hassles.

In order to further investigate the predictive validity of the SDQ teacher version, the degree of coherence between the developmental pathways of SDQ scores and the scores that are indicative of the children’s likability, namely social preference, was examined. Again, we used GMM at the first step to test which model fitted the data best. Table 6 shows that when all fit statistics are taken into consideration these again argue for a model providing three developmental pathways. This can also be seen in Figure 2: One large group scores consistently low on the SDQ total score (81.4%); one group scores high and demonstrates a slight decrease over time (8.7%); and one...
group that starts somewhat lower than the previous group, but shows a small increase over time (9.9%).

At the second step, the developmental pathways were linked to the social preference scores. The results are presented in Table 7. These clearly show that developmental pathways of the SDQ as indicated by teachers, are associated with the extent to which children are liked by their classmates. Hence, the SDQ teacher version demonstrates predictive validity as well. Again, it is noticeable that the two high trajectories hardly differ with regard to social preference scores. The children in the large, stable group are most liked by their classmates.

Normative Data
Finally, normative data are presented for the Dutch population, and for both the parent and teacher version of the SDQ for children aged 4-7. For each child from our sample, we calculated the score on every SDQ subscale and the Total Difficulties score at T1. For each of the five scales, scores vary between 0-10; for Total Difficulties between 0-40. A cumulative percentage equal to or over 95% corresponding to a certain score, means that in the normative sample, 95% of the children acquired lower scores than the child who obtained that particular score, or stated differently, the child belongs to the 5% children exhibiting most problems on that scale. This is referred to as a clinical score. A score corresponding to a cumulative percentage between 90 and 95% is called a subclinical score. Such a score implies that a child belongs to the 10% children exhibiting most problems. Next, we repeated this procedure separately for boys and girls. In Table 8, the scores of the total sample and the scores specified by gender are presented. To facilitate interpretation, we summarized when scores are considered subclinical and clinical as to the five subscales and Total Difficulties score. Generally, it can be stated that the normative scores for the subgroups based on gender, hardly differ from those for the total sample.
In line with our expectations regarding reliability, we found consistently higher omega coefficients than Cronbach’s alpha coefficients. These results mesh with previous studies investigating omega and alpha (Kuijpers et al., 2014; Stone et al., 2013). With the relatively low alpha coefficients being reported previously it has been argued to refrain from using the separate subscales of the SDQ, specifically for the conduct problems and peer problems scales (Niclasen et al., 2013; Stone et al., 2010). However, we showed that these subscales seem to be reliable when an indicator of reliability is employed that takes skewness and difficulties due to limited response categories into account. Therefore, we argue that scores from separate subscales are reliable and thus can be interpreted.

Second, as expected, we were able to confirm the five-factor structure of the SDQ for both the parent and teacher version, which is in line with previous studies employing CFA (Becker et al., 2004; Moriwaki & Kamio, 2014; Niclasen et al., 2013; Palmieri & Smith, 2007; Sanne et al., 2009; Van Leeuwen et al., 2006; Van Roy et al., 2008). Also, we found SDQ scores to be at least configurally invariant across gender, age, and ethnicity for the parent version of the SDQ. On a scalar and metric level SDQ scores were also invariant for gender. These results are largely in line with previous studies (Hill & Hughes, 2007; Niclasen et al., 2013; Palmieri & Smith, 2007; Sanne et al., 2009; Stone et al., 2013). For the teacher version, SDQ scores were also configurally invariant across age, gender and ethnicity. However, for gender, invariance was not established on a scalar and metric level. These inconsistent results are in line with previous studies on the teacher version (Hill & Hughes, 2007; Niclasen et al., 2013; Palmieri & Smith, 2007; Sanne et al., 2009; Van Roy et al., 2008). Also, support for scalar and metric invariance was somewhat inconsistent across time points for both the parent and teacher version regarding age and ethnicity. Further research is warranted on measurement invariance regarding both the parent and teacher version in order to further clarify these inconsistent results.

Regarding predictive validity, we showed inclusion in a risk-group (i.e., the highest SDQ scores in the sample) was predictive of more maladaptive parenting and higher degrees of parenting stress. Also, we found inclusion in a risk-group predictive of lower degrees of being liked, in other words, children who were rated as having more psychosocial problems were less liked by their peers. These results are particularly important for the viability of the SDQ as a screening instrument, as they show that SDQ scores are related to other types of maladjustment over time, attesting to the robustness of the SDQ.

As for criterion validity, we showed that SDQ scores for the parent version were consistently related to maladaptive parenting and parenting stress. Scores on the teacher version were not strongly related to the parenting measures, but were to the sociometric measures. Specifically, the sociometric measures, being liked, disliked and social preference (i.e., the degree to which the child is liked by peers) correlated

### Table 8 Dutch normative data for the parent and teacher version of the SDQ.

<table>
<thead>
<tr>
<th></th>
<th>Parent</th>
<th>Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Boys</td>
</tr>
<tr>
<td>Emotional Symptoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subclinical</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Clinical</td>
<td>5-10</td>
<td>5-10</td>
</tr>
<tr>
<td>Conduct Problems</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Subclinical</td>
<td>4-10</td>
<td>4-10</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Subclinical</td>
<td>7-10</td>
<td>8-10</td>
</tr>
<tr>
<td>Peer problems</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Subclinical</td>
<td>4-10</td>
<td>4-10</td>
</tr>
<tr>
<td>Prosocial behavior</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Subclinical</td>
<td>0-4</td>
<td>0-3</td>
</tr>
<tr>
<td>Total Difficulties</td>
<td>14-16</td>
<td>14-16</td>
</tr>
<tr>
<td>Clinical</td>
<td>17-40</td>
<td>17-40</td>
</tr>
</tbody>
</table>

### Discussion

In the present study, psychometric properties of the parent and teacher version of the SDQ were examined for children aged 4-7 in a large sample. Specifically, omega coefficients and most test-retest indices were adequate, the five-factor structure was confirmed, and indices of criterion validity were adequate. Next, support for measurement invariance was strongest for gender and age, and less so for ethnicity for the parent version. Regarding the teacher version, support for the most stringent type of measurement invariance was not strong across time points, although the less stringent type of measurement invariance was established for age, gender and ethnicity. Further, our results supported the predictive validity of the SDQ. Finally, normative data for the Dutch population were presented. Generally, the SDQ’s psychometric properties can be classified as adequate in this community sample, in young children and with the goal of the SDQ as a screening instrument. Specifically, psychometric properties of the SDQ are dependent on characteristics of the sample and the goal of this study (Hunsely & Mash, 2007). With these notions being made, this study is the first to examine predictive validity of the SDQ while also comprehensively assessing several modern indicators of reliability and validity. As such, this study is an important contribution to the psychometric literature on the SDQ.
substantially with parent and teacher rated scores. These results confirm the criterion validity of SDQ scores for both the parent and teacher version. Moreover, given the stability typically found in sociometric measures (Cillessen & Mayeux, 2004), these measures may be very suited as criterion measures for validation purposes in future studies.

Finally, we presented normative data for children aged 4-7 for the Dutch version of the SDQ enabling researchers and clinicians to interpret SDQ scores as being ‘normal’, ‘subclinical’ or ‘clinical’. When comparing these results to British, Danish and Swedish normative data, our results are largely in line with these studies for both the parent and teacher version (Goodman, 2001; Niclasen et al., 2012; Smedje, Broman, Hetta, & von Knorring, 1999). With the presentation of these norms we facilitate the use of the SDQ as a screening instrument in young children where the potential of prevention and intervention are high. Particularly in these young children this potential may be problems that have probably not yet fully become integrated into the child’s personality. Still, our results show that a small group of children increases in their problem levels. Therefore, targeting such an at-risk group in particular seems a fruitful approach for prevention and intervention.

Some limitations of this study should be noted. First, we did not investigate psychometric properties of the SDQ in a clinical sample and therefore do not know whether our results may be generalized to such a population. As the SDQ is used frequently in clinical practice, either as part of screening at intake or as a routine outcome monitoring instrument (e.g., Van Sonsbeek, Hutschemaekers, Veerman, & Tiernens, 2014), this is an important avenue for future research. Also, although we specifically focused on young children due to limited research concerning this age group, the normative data presented may be quite limited, especially to clinicians. If normative data were established for the complete age range of the SDQ, this would be very useful to clinicians. Second, in this study relatively high attrition levels were found, possibly compromising our results regarding predictive validity and measurement invariance. Therefore, future research into predictive validity and measurement invariance is warranted to replicate our findings. Clinical diagnoses or alternative measures of child adjustment could be included in future studies to enhance the child’s personality. Still, our results show that a small group of children increases in their problem levels. Therefore, targeting such an at-risk group in particular seems a fruitful approach for prevention and intervention.

Some limitations of this study should be noted. First, we did not investigate psychometric properties of the SDQ in a clinical sample and therefore do not know whether our results may be generalized to such a population. As the SDQ is used frequently in clinical practice, either as part of screening at intake or as a routine outcome monitoring instrument (e.g., Van Sonsbeek, Hutschemaekers, Veerman, & Tiernens, 2014), this is an important avenue for future research. Also, although we specifically focused on young children due to limited research concerning this age group, the normative data presented may be quite limited, especially to clinicians. If normative data were established for the complete age range of the SDQ, this would be very useful to clinicians. Second, in this study relatively high attrition levels were found, possibly compromising our results regarding predictive validity and measurement invariance. Therefore, future research into predictive validity and measurement invariance is warranted to replicate our findings. Clinical diagnoses or alternative measures of child adjustment could be included in future studies to examine whether SDQ scores predict maladjustment on these measures. Despite the aforementioned limitations, this study adds to the literature by showing that key aspects of psychometrics, namely reliability, construct validity, measurement invariance and predictive validity were found adequate for the parent and teacher version of the SDQ in this community sample.
Sijsma, K. (2009). On the Use, the Misuse, and Very Limited Usefulness of Cronbach’s Alpha. Psychometrika, 74, 107-120.
Appendix | Measurement invariance analysis

Measurement invariance means that the construct validity is the same for different groups. This includes, for example, the question whether the meaning of the Emotional symptoms scale is the same for boys and girls. In this study, it is assessed whether measurement invariance can be determined with regard to gender, age, and ethnicity. When the measurement invariance of an instrument is established, this allows for meaningful group comparisons (e.g., Meredith, 1993; Vandenbroucke & Lance, 2000; Steenkamp and Baumgartner (1998) distinguished three forms of measurement invariance: configural invariance, metric invariance, and scalar invariance. Configural invariance refers to the question of whether the factors comprise the same set of items for each of the subgroups. Configural invariance is supported if the baseline model fit is acceptable or good. The baseline model is a combination of factor analyses regarding every subgroup. Each subgroup yields a certain fit, that is combined in an aggregated fit across all subgroups. To meet the requirement of configural invariance, this aggregated fit needs to be acceptable or good. Metric invariance means that it is tested whether the loadings are the same across groups. In principal, this is done by equating the loadings across groups, and comparing the fit of this model to the fit of the baseline model. If the fit does not significantly worsen, this supports metric measurement invariance. A more stringent form of measurement invariance concerns the item intercepts. If we consider that the factor model resembles a regression model in which item scores are estimated by the corresponding latent factor, then such a regression model contains intercepts and regression weights. The regression weights are the loadings that, in the previous step, were equated. On top of that, the intercepts are now equated. This is called scalar invariance. The fit of this scalar invariance model is compared to the fit of the previous (metric invariance) model, and it is expected that the fit will not significantly worsen.

For evaluating measurement invariance of scales based on ordinal items (as is the case here), the matter is more complex. Instead of intercepts and factor loadings, we now have to deal with thresholds and factor loadings. Metric and scalar invariance can now only be examined simultaneously (Muthén & Muthén, 1998-2007; Kim & Yoon, 2011). For this purpose, the configural invariance model is compared to the model in which factor loadings and thresholds are equated simultaneously for different groups. To test whether these restriction worsened the model, a chi-square difference test was applied, as indicated on the Mplus website (http://www.statmodel.com/chidiff.html). However, when the chi-square difference test is applied to nested models, as in this case, the test is directly influenced by the sample size. Trivial differences can still be significant to large sample sizes (Schermelleh-Engel, Moosbrugger, & Müller, 2003), and this test is therefore seen as highly sensitive, but not very practical (Cheung & Rensvold, 2002). For that reason, next to this chi-square difference test, we also viewed the fit indices, as is recommended by Cheung and Rensvold (2002). When the chi-square difference test value increases no more than .01 from the baseline model to the restricted model, it can be argued that measurement invariance should not be rejected (cf. Stone et al., 2013).

Parent version

The five factor model was tested, and factor loadings and thresholds were left free to vary. Table 1 shows that the configural invariance of the SDQ parent version is supported at the three measurements with respect to gender, age, and ethnicity (Model 1). The fit is adequate; CFI > .867 and RMSEA < .052. Factor loadings and thresholds were equated to test metric and scalar invariance (Model 2). In Table 1, the fit indices are presented.

T1. Regarding gender, the chi-square difference test results showed no significant difference between Model 1 and Model 2 ($\Delta\chi^2(70) = 86.11, p = .09$). The DIFF test further indicated significant differences regarding age ($\Delta\chi^2(215) = 338.04, p = .000$) and ethnicity ($\Delta\chi^2(70) = 172.45, p = .000$). However, when viewing the differences in the CFI values, it can be seen that these increase no more than .01. Therefore, it can be concluded that there are no substantial differences between the models. Measurement invariance is thus supported as regards gender, age, and ethnicity.

T2. Regarding gender, the DIFF test turned out insignificant ($\Delta\chi^2(70) = 89.52, p = .058$), meaning that Model 1 and 2 do not differ from one another and invariance is supported. It appeared that at T2, there were not enough observations on certain items with respect to age and ethnicity, impeding model identification. Measurement invariance is thus not established regarding age and ethnicity at this time point.

T3. The DIFF test proved significant regarding gender ($\Delta\chi^2(70) = 108.26, p = .002$), and age ($\Delta\chi^2(215) = 302.01, p = .000$). However, the increase in CFI values was no more than .01. It can therefore be concluded that the models do not substantially differ from each other, and measurement invariance is thus supported with regard to these variables. It turned out that at T3, there were not enough observations on certain items as regards ethnicity, hindering model identification. Measurement invariance is thus not established regarding ethnicity at T3.

Teacher version

The five factor model was tested, and factor loadings and thresholds were left free to vary. Table 2 shows that configural invariance of the SDQ teacher version is supported at the three measurements with respect to gender, age, and ethnicity (Model 1). The fit is acceptable; CFI > .900 and RMSEA < .070. Factor loadings and thresholds were equated to test metric and scalar invariance (Model 2). In Table 1, the fit indices are presented.
T1. Regarding ethnicity, the DIFF test proved significant ($\Delta \chi^2(70) = 121.50, p = .000$), but the increase in CFI was no more than .01. Therefore, it can be concluded that Model 1 and 2 do not differ from one another, and measurement invariance is thus supported as regards ethnicity. With respect to gender, the DIFF test turned out significant as well ($\Delta \chi^2(70) = 124.60, p = .000$), and the increase in CFI is somewhat larger than .01. Metric and scalar invariance are thus not completely supported as regards gender. It appeared that at T1, there were not enough observations on certain items with respect to age, impeding model identification. Measurement invariance is thus not established regarding age at this time point.

T2. The DIFF test proved significant regarding gender ($\Delta \chi^2(70) = 138.81, p = .000$), age ($\Delta \chi^2(215) = 346.86, p = .000$), and ethnicity ($\Delta \chi^2(70) = 116.76, p = .000$). The increase in CFI was larger than .01. Metric and scalar invariance are thus not supported as regards these variables.

T3. Regarding age and ethnicity, the DIFF test turned out significant ($\Delta \chi^2(215) = 442.02, p = .000$; and $\Delta \chi^2(70) = 129.86, p = .000$, respectively). The increase in CFI, however, appeared to be no more than .01. Therefore, it can be concluded that there are no substantial differences between Model 1 and 2. Measurement invariance is thus supported as regards age and ethnicity. The DIFF test proved significant for gender as well ($\Delta \chi^2(70) = 125.72, p = .000$), yet the increase in CFI was somewhat larger than .01. Therefore, metric and scalar invariance are not completely supported with respect to gender.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Fit indices of the SDQ parent version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Factor loadings and thresholds</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
</tr>
</tbody>
</table>

| T2 | Free to vary | Gender | 1116.44 | 530 | .000 | .897 | .047 |
| Equal | | 1121.78 | 600 | .000 | .908 | .042 |

| T3 | Free to vary | Gender | 999.62 | 530 | .000 | .927 | .045 |
| Equal | | 1035.68 | 600 | .000 | .932 | .041 |
| Free to vary | Age | 1665.64 | 1060 | .000 | .908 | .051 |
| Equal | | 1837.43 | 1275 | .000 | .915 | .045 |

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Fit indices of the SDQ teacher version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>Factor loadings and thresholds</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
</tr>
<tr>
<td>1</td>
<td>Free to vary</td>
</tr>
<tr>
<td>2</td>
<td>Equal</td>
</tr>
</tbody>
</table>

| T2 | Free to vary | Gender | 2692.35 | 530 | .000 | .928 | .068 |
| Equal | | 2409.60 | 600 | .000 | .940 | .058 |
| Free to vary | Age | 3520.41 | 1060 | .000 | .928 | .072 |
| Equal | | 3285.75 | 1275 | .000 | .941 | .060 |

| T3 | Free to vary | Ethnicity | 2838.26 | 530 | .000 | .936 | .066 |
| Equal | | 2479.87 | 600 | .000 | .948 | .056 |

| Free to vary | Gender | 2244.52 | 530 | .000 | .926 | .066 |
| Equal | | 2060.66 | 600 | .000 | .937 | .057 |
| Free to vary | Age | 2918.37 | 1060 | .000 | .930 | .069 |
| Equal | | 2944.26 | 1275 | .000 | .938 | .060 |

| Free to vary | Ethnicity | 2078.26 | 530 | .000 | .940 | .062 |
| Equal | | 1898.12 | 600 | .000 | .950 | .054 |
Chapter 5

Relations Among Internalizing and Externalizing Problems in Early Childhood
Introduction

Internalizing and externalizing problems represent the two major psychopathologies of early childhood, are closely related and tend to co-occur (Achenbach, Howell, Quay, & Conners, 1991; Lilienfeld, 2003; Oland & Shaw, 2005). Although research has begun to unravel how and why these problems are associated, it remains unclear whether third variables may explain the relations between internalizing and externalizing problems. As such, it is deemed important to simultaneously examine a directional model and a third variable model of childhood psychopathology (cf. Lee & Bukowski, 2012; Mathiesen, Sanson, Stoolmiller, & Karevold, 2009). Therefore, the current study investigates whether internalizing and externalizing problems are reciprocally or unidirectionally related over time. Further, it tests whether relations between internalizing and externalizing problems might be explained by third variables.

The interrelatedness of internalizing and externalizing problems in childhood and adolescence has largely been studied from a directional model perspective. Studies starting from this perspective have addressed whether internalizing problems precede externalizing problems (Bittner, Egger, Erkanli, Costello, Foley, & Angold, 2007; Ialongo, Edelsohn, Werthamer-Larsson, Crockett, & Kellam, 1996; Last, Perrin, Hersen, & Kazdin, 1996; Ritakallio, Koivisto, von der Pahlen, Pelkonen, Marttunen, & Kaltiala-Heino, 2008; Vitaro, Brendgen, & Tremblay, 2002), which is in line with the acting out hypothesis that depressive symptoms lead to acting out behaviors (Glaser, 1967), or the hypothesis that anxiety underlies aggression (Patterson & Capaldi, 1990), and indeed found externalizing problems to precede internalizing problems (Boylan, Vaillancourt, Boyle, & Szatmari, 2007; Boylan, Vaillancourt, & Szatmari, 2007; Boylan, Vaillancourt, & Szatmari, 2012; Burke, Loeber, Lahey, & Rathouz, 2005; Copeland, Shanahan, Costello, & Angold, 2005; Lahey, Loeber, Burke, Rathouz, & McBurnett, 2002; Speltz, McClellan, DeKlyen, & Jones, 1999). Finally, studies have addressed whether change in one cluster of problems is associated with failures in social situations that in turn lead to anxiety and depression (Patterson & Capaldi, 1990), and indeed found externalizing problems to precede internalizing problems (Boylan, Vaillancourt, Boyle, & Szatmari, 2007; Boylan, Vaillancourt, & Szatmari, 2012; Burke, Loeber, Lahey, & Rathouz, 2005; Copeland, Shanahan, Costello, & Angold, 2005; Hoglund & Leadbeater, 2004; Lahey, Loeber, Burke, Rathouz, & McBurnett, 2002; Speltz, McClellan, DeKlyen, & Jones, 1999). Finally, studies have addressed whether change in one cluster of problems is associated with changes in the other and act as risk factor to the other (Gilliom & Shaw, 2004; Keiley, Bates, Dodge, & Pettit, 2000; Lee & Bukowski, 2012; Mesman, Bongers, & Koot, 2001). While these studies provide evidence for direct relations between internalizing and externalizing problems, there are also studies examining the role of third variables in explaining this relationship (Fergusson, Lyonskey, & Horwood, 1996; Weiss, Süsser, & Catron, 1998). Third variables may act as latent or indirectly observed tendencies to develop a disorder (Krueger & Markon, 2006). Consequently, third variables could be described as those factors that are related to multiple disorders, such as internalizing and externalizing problems, and are hypothesized to underlie both problems.
As stated by many researchers, studies on co-variation of disorders are highly complex and subject to problems on various levels (heterogeneity of terminology concerning ‘comorbidity’, sampling or referral bias, informant bias, variability in diagnostic and analytic procedures, taxonomic problems and symptom overlap (Angold, Costello & Erkanli, 1999; Krueger & Markon, 2006). Although recently there is increasing evidence that covariation of disorders is more than a conceptual artifact or methodological nuisance, developmental pathways are far from being completely unraveled. This study adds to the plethora of research in various ways. First, most studies have focused on either the directional or the third variables model, and have not evaluated these models simultaneously. Second, few studies have studied multiple liability factors and corrected for spurious relations. This study focuses on several contextual liability factors known to be related to internalizing and externalizing problems. Second, this study employs rigorous modeling techniques such as cross-lagged modeling. Fourth, we focus on young children where relations between internalizing and externalizing may be different than for older children or adolescents. According to transactional ecological models of psychopathology, the interplay between biological, psychological and social systems contribute to the development of internalizing and externalizing problems (Bronfenbrenner & Evans, 2000; Cicchetti & Toth, 1998). Specifically, factors from social systems such as parents and peers may underlie both internalizing and externalizing problems as they are proximal to the child and hypothesized to exert great influence during early childhood (Cicchetti & Toth, 1998; Ford, Collishaw, Meltzer, & Goodman, 2007). As such, the covariation among internalizing and externalizing problems may be in part explained by common factors in the social system wherein the child is developing. Indeed, inadequate parenting, parenting stress and maternal mental health have been strongly linked to internalizing and externalizing problems (e.g., Arnold, O’Leary, Wolff, & Acker, 1993; Barry, Dunlap, Cotten, Lochman, & Wells, 2005; Beardslee, Versage, & Gladstone, 1998; Gross, Shaw, & Moilanen, 2008; Prinzie, Onghena, Hellinckx, Ghesquiere, & Colpin, 2003; Rodriguez, 2011). Furthermore, when a child is disliked by more children than that it is liked, indicated by low social preference, effects on internalizing and externalizing problems have repeatedly been reported, possibly by affecting the child’s self-perceptions (e.g., Laird, Jordan, Dodge, Pettit, & Bates, 2001; Gooren, Van Lier, Stegge, Terwogt, & Koot, 2011). Thus, inadequate parenting, parenting stress, maternal mental health and low social preference are hypothesized to act as third variables, thereby explaining a possible spurious relation between internalizing and externalizing problems.

Some empirical evidence supports the premise that third variables partially explain the relations between internalizing and externalizing problems (Gjone & Stevenson, 1997; Kessler, Petukhova, & Zaslavsky, 2011). In a sample of children followed from 18 months to 4.5 years, internalizing and externalizing problems were moderately related. Yet, when controlling for family stress, partner support and child emotionality, this relation became substantially weaker (Mathiesen et al., 2009). Similarly, in boys aged 10-13, but not girls, reciprocal effects of externalizing and internalizing problems over time became smaller when including parental violence to the model (Lee & Bukowski, 2012). However, some studies do not support that third variables explain these results, but show that relations hold while controlling for these variables. In a study that followed preschoolers through adolescence, predictive paths from externalizing problems to internalizing problems were found, but these were not altered by the inclusion of the third variable social problems (Mesman et al., 2001). Again, a study following preschoolers through adolescence found externalizing problems to impact academic competence which in turn impacted internalizing problems. These associations held while controlling for gender, socioeconomic status, early caregiving and cognitive ability (Burt & Roisman, 2010). Children followed from early childhood to adolescence were found to have more problems when their social competence was lower at preschool, when controlling for intelligence, maternal education and social desirability (Bornstein, Hahn, & Haynes, 2010). Clearly, more research is needed in order to examine how internalizing and externalizing problems are related over time. These studies have not been conducted in early childhood, where effects of common environmental influences on co-occurring internalizing and externalizing problems were found to be greatest (Gjone & Stevenson, 1997). Further, these studies both employed latent growth curve modeling, a person-centered approach. Although person-centered analyses are important, variable-centered analyses are advised when investigating associations between variables, and when investigating the relative contribution that a predictor variable (e.g., externalizing problems) makes to an outcome (e.g. internalizing problems) (Laursen & Holf, 2006). Also, the risk factors included in these studies did not take parenting into account, which is known to be an important factor in the development of child psychopathology (e.g. Dishion & Patterson, 2006). Finally, the former studies did not emphasize whether the two problem clusters predict each other’s onset or stability, nor whether the third variables impact onset or stability specifically. It seems essential to establish which third variables impact onset versus stability. As such, it would be possible to distinguish between factors that set the stage for problem behavior versus factors that affect symptoms when the problem behavior is already present.

In sum, the purpose of the present longitudinal study is to test both a directional and third variables model of internalizing and externalizing problems in early childhood in a large sample. First, we examine whether parent-reported internalizing and externalizing problems are related reciprocally or unidirectionally. We expect that externalizing problems are strongly related to subsequent internalizing problems, and vice versa. Second, we will identify third variables that are related to internalizing
and externalizing problems over time. Specifically, we will simultaneously examine the associations of inadequate parenting, parenting stress, maternal mental health, and social preference with both internalizing and externalizing problems, while controlling for the other problem cluster. It is expected that, parenting stress, inadequate parenting, maternal mental health, and social preference act as third variables for both internalizing and externalizing problems. Further, we investigate whether internalizing and externalizing problems are related to each others’ onset and stability and to what extent third variables contribute to these associations. It is expected that internalizing problems are related to stability in externalizing problems, and vice versa. Furthermore, all third variables are expected to be related to onset and stability of internalizing and externalizing problems.

Method

Sample and Procedure

Mothers of children aged 4-7 from 29 primary schools throughout the Netherlands were recruited for the Dutch “Child in Sight (Kind in Zicht)” study, of whom 1,339 mothers filled in questionnaires for their children (M age = 5.08, SD = 1.25, 50.1% boys) in the first assessment (T1). In a subsequent assessment one year later (T2) the participation rate was 67%, with 95 parents who did not participate in the baseline assessment. Due to use of Structural Equation Modeling, wherein missing cases are accounted for, our final number of participants is 1,434. At baseline, mothers had a mean age of 36.61 (SD = 4.41), the majority was of Dutch origin (92.4%) and were part of a two-parent household (89.1%). Most mothers, 44.6%, were highly educated with a college or university degree, 37.8% finished vocational education and 13.7% finished a low level of Dutch secondary school. 3.9% finished a different form of education. A logistic regression analysis showed that families who completed two waves (n=817) did not differ from the dropouts (n=522) in child age, gender, maternal education. A logistic regression analysis showed that families who completed two waves (n=817) did not differ from the dropouts (n=522) in child age, gender, maternal education. A logistic regression analysis showed that families who completed two waves (n=817) did not differ from the dropouts (n=522) in child age, gender, maternal education.

We used data of two annual waves of Kind in Zicht, a large cohort study of Dutch educational level, internalizing, and externalizing problems. Each scale contains five items and parents rated their child’s behavior on a 3-point scale ranging from 0 (not true) to 2 (certainly true). The scoring procedures used in this study are available online at www.sdqinfo.com. As scale distributions of the SDQ are skewed, alternative indicators of reliability based on Structural Equation Modeling – are known as Jöreskog rho or McDonald’s Omega_H, were used to assess reliability (Jöreskog, 1971; McDonald, 1978,1999; Revelle & Zinbarg, 2009; Stone, Otten, Ringlever, Hiemstra, Engels, Vermulst & Janssens, 2013a). Omega (ω) values were .79 and .80 at T1 and T2 for the emotional symptoms scale, and .71 and .75 at T1 and T2 for the conduct problems scale.

Inadequate Parenting. The Parenting Scale was used at the first wave and asks parents to rate 30 short parenting situations on a 7-point scale (Arnold et al., 1993). Sample items include “When I want my child to stop doing something I firmly tell my child to stop / I coax or beg my child to stop” and “When I’m upset or under stress I am picky and on my child’s back / I am no more picky than usual”. Inadequate parenting behavior is divided across three subscales: permissiveness, restrictiveness, and social preference with both internalizing and externalizing problems.

Measures

Internalizing and Externalizing Problems. The Dutch parent version of a screening questionnaire for psychopathology, the Strengths and Difficulties Questionnaire was used at both waves to assess internalizing and externalizing problems (SDQ; Widenfert, Goedhart, Treffers, & Goodman, 2003). The SDQ has been shown reliable and valid for use in a community sample (see for a review see for a review Stone, Otten, Engels, Vermulst, & Janssens, 2010). The subscale emotional symptoms (e.g., many worries, often seems worried) was used to measure internalizing problems. Each conduct problems scale (e.g., often lies or cheats) was used to measure externalizing problems. Each scale contains five items and parents rated their child’s behavior on a 3-point scale ranging from 0 (not true) to 2 (certainly true). The scoring procedures used in this study are available online at www.sdqinfo.com. As scale distributions of the SDQ are skewed, alternative indicators of reliability based on Structural Equation Modeling – are known as Jöreskog rho or McDonald’s Omega_H, were used to assess reliability (Jöreskog, 1971; McDonald, 1978,1999; Revelle & Zinbarg, 2009; Stone, Otten, Ringlever, Hiemstra, Engels, Vermulst & Janssens, 2013a). Omega (ω) values were .79 and .80 at T1 and T2 for the emotional symptoms scale, and .71 and .75 at T1 and T2 for the conduct problems scale.

Parenting Stress. At the first wave mothers rated the frequency of daily hassles with their child over the past 6 months (Parenting Daily Hassles: PDH; Crnic & Booth, 1991). The questionnaire consists of 20 events of which the parent has to rate how often (seldom, sometimes, often, constantly) they occur. Sample items include
“Continually cleaning up messes of food and toys” and “The kids demand that you entertain them or play with them”. Psychometric properties of the PDH have been found adequate (Crnic & Booth, 1991; Rispens, Hermanns, & Meeus, 1996). Cronbach’s alpha was .77.

Mental Health. The degree of mental health of the mothers during the past 4 weeks was measured at the first wave with a short version of the General Health Questionnaire (GHQ; Hardy, Shapiro, Haynes & Rick, 1999). Mothers rated their mental health via 12 questions (e.g., did you lose confidence in yourself? did you feel able to make decisions?) on a 4-point scale. The summed items yield a total score, with higher scores indicating diminished mental health. Research into reliability and validity indicates that the GHQ has adequate psychometric properties (Koeter & Ormel, 1991). Cronbach’s alpha was .89.

Social Preference. During individual interviews, children were shown a photograph of their classmates. A trained research assistant pointed out a child on the photograph and asked the child whether(s)he knew who this child was, ensuring familiarity, and was then asked whether(s)he liked, disliked the child or thought neutral of him/her. To increase comprehension and ease shy children, the child could respond verbally or by pointing to three fluffy smileys, with either a happy, sad or neutral expression. This procedure was repeated until the child gave a nomination about every child in the class. The order of asking questions about children in the photograph was counterbalanced, such that the interviewer started either at the upper left, upper right, lower left or lower right corner of the photograph. Unlimited nominations (like, dislike, neutral) were used, because these tend to spread more evenly among children in a class than limited nominations (i.e., fewer children receive a raw nomination score of zero). The total least-liked nomination was subtracted from the total most-liked nomination to obtain a measure of social preference. Social preference scores were then standardized within each classroom (cf. Coie, Dodge, & Coppotelli, 1982).

Strategy for analysis
First, means, standard deviations and bivariate correlations of all study variables were calculated. Second, to evaluate the associations of internalizing and externalizing problems over time, we tested a latent cross-lagged path model while controlling for gender and age, and subsequently including the third variables inadequate parenting, parenting stress, maternal mental health and social preference in the model, using MPLUS version 5 (Muthén & Muthén, 1998-2007; see Figure 1 for the conceptual model). Internalizing and externalizing problems are latent variables measured by five items each. The items have 3-points response scales and are mostly very skewed. For this type of items (denoted as ordered categorical in Mplus) we used the Weighted Least Square estimator with Mean- and Variance-adjusted chi-square test statistic (WLSMV). Asparouhov and Muthén (2010) have investigated the consistency of the WLS(MV)-estimator under various missing data assumptions and concluded that using all available pairwise information in the data produced unbiased and efficient estimates for the parameters to be estimated. Third, we tested whether the model parameters differed for children with no internalizing (n = 1,243, score 0-3) or externalizing problems at T1 (n = 1,200, score 0-2) (non-cases) and children with internalizing (n =191, score 4-10) or externalizing problems (n =234, score 3-10) in the clinical range (cases) at T1. Continuous internalizing and externalizing scores were classified into ‘non-cases’ and ‘cases’ scores based on the SDQ’s scoring procedures, at T1 and T2, such that approximately 90 % of children are classified as non-cases. Goodman (2001) examined the clinical validity of this scoring method is extensively and found adequate clinical validity for the parent version of the SDQ. We tested four models with one including non-cases of internalizing problems at T1 and predicting onset of internalizing ‘caseness’ at T2 (e.g., predicting a score of 0 = no clinical level internalizing problems, versus 1 = clinical level internalizing problems present), and one including cases of internalizing problems at T1 and predicting stability in internalizing problems at T2. The third model includes non-cases of externalizing problems at T1 and predicts onset of externalizing problems at T2. The fourth model includes cases of externalizing problems at T1 and predicts stability in externalizing problems at T2. Because the use of latent constructs would lead to too many parameters to be estimated for the sample size that we had, we used observed variables for the onset and stability analyses. In all of these models again we controlled for gender and age.
for gender and age, and the third variables were included to test whether third variables accounted for the variance in the possible association between internalizing and externalizing problems. Also, we controlled for concurrent associations between internalizing and externalizing problems at T2. As our dependent variable in the onset and stability analyses is categorical, we used the WLSMV-estimator.

As children are in the same classes, our data may be nested such that children from the same classes may share common behaviours (i.e., clustering). Therefore, we corrected for this in our analyses using the TYPE is COMPLEX command in Mplus.

Results

Descriptive statistics
As expected, internalizing and externalizing problems were significantly related (Table 1). Further, internalizing problems at T1 were strongly correlated with internalizing problems at T2, indicating high stability of these problems. A similar association was found for externalizing problems, again indicating stability of these problems in young children. Internalizing problems at T1 were most strongly related to parenting stress at T1, and internalizing problems at T2 were most strongly related to parenting stress, and maternal mental health, while externalizing problems at T1 and T2 were most strongly associated with inadequate parenting, parenting stress and social preference. At both waves, more internalizing than externalizing problems were reported. Finally, on average, internalizing problems increased from T1 to T2, whereas externalizing problems decreased.

Model findings
Fit statistics for the model investigating the relations between internalizing and externalizing problems were satisfactory ($\chi^2(186) = 349.43, p < 0.00; CFI = 0.952; RMSEA = 0.028 (CI = 0.023--0.033); TLI = 0.941$). Standardized estimates are presented in Figure 2. Factor loadings were as follows for internalizing problems items at T1 (.42, .78, .77, .70, .72) and T2 (.42, .78, .71, .71, .74), and externalizing problems items at T1 (.76, .48, .74, .52, .37) and T2 (.72, .46, .62, .68, .50). Internalizing problems at T1 predicted subsequent internalizing problems, and externalizing problems at T1 predicted externalizing problems at T2. No cross-lagged associations were found. Gender was negatively associated with externalizing problems at T1 and T2. This means that mothers reported more externalizing problems for boys than for girls. Age was positively related to internalizing problems at T1 and T2. This means that mothers reported more internalizing problems for boys than for girls.

| Table 1 Correlations between all study variables (N=1,434) |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | 1               | 2               | 3               | 4               | 5               | 6               | 7               | 8               |
| 1 Internalizing T1 | 1.62 (1.83)     | -               |                |                |                |                |                |                |
| 2 Externalizing T1 | 1.26 (1.42)     | .22**           | -               |                |                |                |                |                |
| 3 Inadequate parenting T1 | 2.72 (2.00) | .23**           | .13**           | -               |                |                |                |                |
| 4 Parenting stress T1 | 1.49 (1.28) | .23**           | .34**           | .26**           | -               |                |                |                |
| 5 Mental health T1 | 1.44 (2.50)     | .10**           | .11**           | .18**           | .18**           | -               |                |                |
| 6 Social preference T1 | .06 (.95) | -.07            | -.23**          | -.09**          | -.13**          | -.01           | -.01           | -.19**          |
| 7 Internalizing T2 | 1.71 (1.89)     | .60**           | .18**           | .10**           | .22**           | -.01           | -.01           | -.19**          |
| 8 Externalizing T2 | 1.15 (1.37)     | .15**           | .15**           | .16**           | .32**           | .10**          | .10**          | .27**           |

Note. Means with similar superscripts (a, b, c, d) are significantly different (p < 0.05). Paired t-tests were used.

*p < .05, **p < .01, ***p < .001.
When including the third variables in the model, the fit statistics were satisfactory ($\chi^2(250) = 452.65, p < 0.00; CFI = 0.944; RMSEA = 0.027 (CI = 0.023–0.031); TLI = 0.927$). Standardized estimates are presented in Figure 3. Factor loadings were as follows for internalizing problems items at T1 (.43, .77, .69, .72) and T2 (.43, .78, .72, .70, .74), and externalizing problems items at T1 (.73, .53, .73, .52, .37) and T2 (.71, .51, .62, .67, .43). Internalizing problems at T2 were predicted by maternal general health, such that the more health related problems mothers experienced at T1 the more internalizing problems their children showed one year later. For externalizing problems, parenting stress was positively related to externalizing problems, which means that the more parenting stress mothers experienced at T1, the more externalizing problems their children had one year later. None of the other variables predicted internalizing and externalizing problems. No cross-lagged paths were found from internalizing to externalizing problems and vice versa. Gender was negatively associated with externalizing problems at T2. This means that mothers reported more externalizing problems for boys than for girls. Age was positively related to internalizing problems at T1, indicating that mothers reported more internalizing problems for older children.

Furthermore, inadequate parenting and parenting stress were positively associated to externalizing problems concurrently, such that the more inadequate parenting and parenting stress were reported, the more externalizing problems. Also, social preference at T1 was negatively related to externalizing problems at T1. The lower the child’s social status, the more externalizing problems were reported. These relations were absent longitudinally. Parenting stress was related to internalizing problems concurrently, such that the more parenting stress was reported, the more internalizing problems children showed. Again, this relation was absent longitudinally.

### Onset and stability

The standardized estimates of the models for onset and stability are presented in Table 2. These models were saturated; therefore no fit statistics are given. Externalizing problems of children at T1 were related to onset and stability in internalizing problems at T2 for children with clinical internalizing scores at T1. When we controlled for third variables, the relation of externalizing problems with onset of clinical level internalizing problems disappeared. This means that externalizing problems are related to stability of already existing internalizing problems over time quite robustly, while this does not hold for onset of clinical level internalizing problems. Regarding third variables, maternal mental health was related to both onset and stability of clinical level internalizing problems. This indicates that the more mental health related problems

---

**Figure 2** Latent cross-lagged model without controlling for third variables. For clarity, factor loadings are given in the text. Numbers after variable names refer to data waves.

**Figure 3** Latent cross-lagged model, controlling for third variables. For clarity, factor loadings are given in the text. Variable abbreviations include inadequate parenting (IP), parenting stress (PS), mental health (MH), social preference (SP). Numbers after variable names refer to data waves.
mothers experienced, the more internalizing problems they reported at T2 for children who were rated as non-clinical in their internalizing problems at T1 and for children who were rated as clinical in their internalizing problems at T1. Inadequate parenting and parenting stress were related to internalizing problems at T2 for children who were rated as non-clinical in their internalizing problems at T1, but were not related to stability of internalizing problems at T2 for children who were already rated as having clinical internalizing problems at T1. This indicates that inadequate parenting and parenting stress were related to the onset of internalizing problems, but not on the stability of these problems over time.

Internalizing problems were related to onset of externalizing problems, but this relation disappeared when we controlled for third variables. Further, internalizing problems were not related to stability of clinical level externalizing problems. As for third variables, only parenting stress was related to the stability of clinical level externalizing problems. This indicates that the more parenting stress mothers experienced, the more externalizing problems they reported at T2 for children who were rated as clinical in their externalizing problems at T1.

Discussion

The current study investigated whether parent reported internalizing and externalizing problems are related unidirectionally or reciprocally in early childhood and whether third variables may explain these relations. Our longitudinal design and large sample size permitted us to test these questions using a rigorous analytical approach. Therefore, the results of this study add to the body of literature investigating relations among internalizing and externalizing problems. We found strong relative stability of internalizing and externalizing problems over time. Further, evidence for these problem clusters reciprocally influencing each other seemed to depend on severity of the problems. When problem level was not taken into account, externalizing problems were not related to subsequent internalizing problems. Internalizing problems did not have any relation with externalizing problems one year later in the total sample. Still, when classifying the internalizing and externalizing scores as clinical and non-clinical based on the norm cut-off we employed, externalizing problems at baseline were related to onset of internalizing problems one year later and to stability of already existing internalizing problems at baseline. The effects regarding stability, but not onset, remained strong when we controlled for third variables, indicating that externalizing problems are substantially related to existing clinical internalizing problems. Internalizing problems were related to onset of subsequent externalizing problems, but not to stability of externalizing problems. Moreover, this relation diminished when third variables were taken into account.

From these results, we may conclude that externalizing problems have a robust association with subsequent clinically elevated internalizing problems, even when controlling for third variables. As such, these results are partly in favor of a directional model and a third variables model (Fergusson et al., 1996; Patterson & Capaldi, 1990; Weiss et al., 1998). Furthermore, these results also point out that we did not find evidence for the ‘acting out’ or ‘anxiety underlying aggression’ hypotheses from directional models in the current study as internalizing problems were not related to subsequent externalizing problems when controlling for third variables (Glaser, 1967; Granič, 2012).

According to one of the directional models (i.e., the failure model), conduct problems lead to failures in social situations that in turn lead to anxiety and depression (Patterson & Capaldi, 1990). According to this model one would expect that externalizing problems at baseline predict ‘new cases’ of internalizing problems one year later (cf. Lahey et al., 2002). Our results only partly support this, as this relation became non-significant when including parent and peer factors. We did find that concurrent externalizing problems were related to subsequent clinically elevated internalizing problems for children already scoring in the 10% highest range at baseline, even when controlling for inadequate parenting, parenting stress, maternal

### Table 2 Standardized estimates for onset and stability

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Internalizing</th>
<th>Externalizing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Onset T2</td>
<td>Stability T2</td>
</tr>
<tr>
<td></td>
<td>(n=1,243)</td>
<td>(n=191)</td>
</tr>
<tr>
<td>Externalizing T1</td>
<td>.15**</td>
<td>.34**</td>
</tr>
<tr>
<td>Internalizing T1</td>
<td>-</td>
<td>.15*</td>
</tr>
<tr>
<td>Total R²</td>
<td>.04</td>
<td>.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2 Predictors (β)</th>
<th>Internalizing</th>
<th>Externalizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate parenting</td>
<td>.16*</td>
<td>-.14</td>
</tr>
<tr>
<td>Parenting stress</td>
<td>.14*</td>
<td>.08</td>
</tr>
<tr>
<td>Mental health</td>
<td>.14**</td>
<td>.24**</td>
</tr>
<tr>
<td>Social preference</td>
<td>.05</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note: *p < .05, **p < .01
mental health and social preference. Conclusively, the present findings suggest that 
externalizing problems in early childhood may better be viewed as a strong 
maintaining factor in clinical internalizing problems and less likely to be a precipitating 
factor as the Patterson and Capaldi model would suggest (Lahey et al., 2002). As our 
study was the first to employ cross-lagged modeling with continuous and categorical 
variables, we contributed to the large body of literature by showing that the role of 
externalizing problems in the development of internalizing problems seems to 
depends on its severity.

Regarding the third variables model, when observing the stronger relations of 
externalizing with internalizing problems in the group with high internalizing levels at 
baseline, inclusion of third variables did not decrease this relation. On the other hand, 
results regarding onset do offer some support for the third variables model, in that 
relations of internalizing problems with onset of externalizing problems, and vice 
versa became weaker and nonsignificant when including third variables. These 
results thus partly support the third variables model (Fergusson et al., 1996; Weiss et 
al., 1998) and are partly in line with former studies, which also reported a decrease of 
reciprocal influences of internalizing and externalizing, although not a complete 
disappearance when including third variables (Lee & Bukowski, 2012; Mathiesen et 
al., 2009). Possibly, relatively stable third variables such as genetic liability or 
temperament, could potentially account for the association of externalizing problems 
with elevated internalizing problems (Keiley et al., 2003; O’Connor, McGuire, Reiss, 
Hetherington, & Plomin, 1998).

Still, we did find that some of the included variables were specifically related to 
either internalizing or externalizing problems, and we were able to distinguish which 
variables are related to onset of problem behavior and which affect symptoms in 
children with elevated levels of problems at baseline. First, parenting stress was 
related to subsequent externalizing problems in the total sample and onset of clinical 
internalizing problems. These results are in line with studies showing strong 
correlated links of parenting stress with internalizing problems (e.g., Rodriguez, 
2011), longitudinal effects of parenting stress on externalizing problems (e.g., Benzies, 
Harrison, & Magill-Evans, 2004), and highlight the salience of parenting stress 
(Deater-Deckard, 2004). Parenting stress also was related to stability of clinical 
externalizing problems but not of clinical internalizing problems. This suggests that 
parenting stress might be a maintaining factor for clinical externalizing problems, 
which may be explained by the lack of consistent parenting associated with parenting 

Second, maternal mental health was related to subsequent internalizing problems 
in the total sample and onset and stability of clinically elevated internalizing problems. 
Possibly, maternal mental health can be seen as a broad-band specific feature 
(Weiss et al., 1998), which relates primarily to the broad-band concept of internalizing 
problems specifically but not to externalizing problems. However, this finding is in 
contradiction to studies that have reported a link between maternal mental health and 
externalizing problems (Fanti & Henrich, 2010; Gross, Shaw, & Moilanen, 2008). 
Although these studies did not control for parenting stress and inadequate parenting 
which are known to be related to maternal mental health (Deater-Deckard, 2004), 
more research is needed to draw conclusions regarding the role of maternal mental 
health in relation to internalizing and externalizing problems.

Third, social preference was related to externalizing problems in concurrently, 
but not longitudinally and not to onset and stability of clinical externalizing problems. 
These results coincide with research indicating that peer rejection is related to 
externalizing problems (Keiley et al., 2003). Exclusion by peers is proposed to lead to, 
or increase acting out behaviors (Van Lier & Koot, 2010; Loeber & Keenan, 1994), 
although our analyses regarding onset and stability do not support this reasoning. 
The reverse might also be true; children who show externalizing problems may be 
particularly prone to become excluded (Hammen, 2006; Panak & Garber, 1992).

Fourth, inadequate parenting was not related to internalizing and externalizing 
problems in the total sample, although concurrent relations were present. Apparently, 
the effects of maternal mental health and parenting stress are stronger than those of 
inadequate parenting when simultaneously evaluated. Inadequate parenting was 
related to onset of internalizing problems, and marginally to onset of externalizing 
problems. As such, inadequate parenting may be a risk factor for onset of multiple 
problems. It is possible that parenting constructs that are more disturbing or intrusive 
in nature may relate more strongly to both problem clusters, such as harsh punishment 
and psychological control (Keiley et al., 2003; Rather, Fite, & Gaertner, 2011, Stone, 
Otten, Janssens, Soenens, Kuntsche, & Engels, 2013b).

The present study is not without limitations. First, although our study was 
longitudinal, it included only two time-points. With more assessment waves it is possible 
to examine whether internalizing and externalizing problems are related to each other 
over a longer course in childhood and adolescence. Also, this would enable more 
fine-grained analyses, such as the identification of trajectories of internalizing and 
externalizing problems or latent growth curves (Nagin, 1999; Willett & Sayer, 1994). It 
is possible that the observed relations between internalizing and externalizing 
problems differ dependent on its specific trajectory. For example, reciprocal 
influences of externalizing on internalizing problems may be stronger in children 
following an increasing internalizing and externalizing trajectory. In order to fully 
derstand how and why internalizing and externalizing problems are related, it 
should also be investigated whether these relations hold when including common 
risk factors. Future studies should include three or more time points, enabling them to 
apply sensitive statistical techniques, such as growth curve modeling in order to 
test developmental change and to test several developmental paths (e.g., linear,
INTERNALIZING AND EXTERNALIZING PROBLEMS

References


Abstract

A key factor identified in friendship formation and stability is similarity. Homophily of externalizing problems has been reported frequently, but less attention has been directed at homophily of internalizing problems. Whether young children who are friends resemble each other in their internalizing problems is thus largely unknown. In order to increase understanding of the social risk factors implicated in the etiology of internalizing problems, it is important to establish whether internalizing problems cluster in friendships. The present study examines homophily of internalizing problems while controlling for externalizing problems in a sample of children aged 4-8. A community sample of 1,584 children or 792 dyads with a mean age of 5.85 (SD = 1.26) participated. Of these friendship dyads, 557 were reciprocal and 235 unilateral friends. Internalizing and externalizing problems were reported by teachers. Intraclass correlations were calculated to test for similarities between reciprocal and unilateral friends. Support was found for the homophily hypothesis, in that reciprocal, but not unilateral friends, were substantially similar in their internalizing problems. In reciprocal friendships, we found substantial similarity. From these results, we might conclude that homophily is already present in friendships of young children regarding internalizing problems. Theoretical and practical implications of these findings are discussed.

Introduction

Internalizing problems belong to the most common forms of child psychopathology (Canino et al., 2004; Egger & Angold, 2006; Ford, Goodman, & Meltzer, 2003; Ollendick & King, 1994) and are characterized by disturbances in emotions and moods, which most often result in anxiety and depressive disorders (Ollendick & King, 1994; Zahn-Waxler, Klimes-Dougan, & Slattery, 2000). Internalizing problems refer to those feelings of distress that are directed inwards and signify a core disturbance in emotions and moods, such as sorrow, guilt, fear, and worry (Zahn-Waxler et al., 2000). As prevalence rates of internalizing problems are highest in adolescence and adulthood (e.g., Cole et al., 2002; Keenan-Miller, Hamm, & Brennan, 2007; Twenge & Nolen-Hoecksema, 2002), research has focused primarily on adolescent and adult internalizing problems (Zahn-Waxler et al., 2000). However, evidence suggests that the foundation for the development of internalizing problems is laid in childhood (Hankin, Badanes, Abela, & Wamnura, 2010; Luby, St. Belden, Tandon, & Spitznagel, 2009; Zahn-Waxler et al., 2000). So, to understand the dramatic increase in internalizing problems during adolescence, investigating factors associated with childhood internalizing problems is essential (Cicchetti & Cohen, 2006; Luby, 2010; Ollendick & King, 1994; Zahn-Waxler et al., 2000).

Specifically, social risk factors are significantly related to adolescent internalizing problems that internalizing problems tend to cluster in friendships (Brendgen, Lamarche, Warner, & Vitaro, 2010; Giletta et al., 2011; Hogue & Steinberg, 1995; Prinstein, 2007; Stevens & Prinstein, 2005; Van Zaik, Van Zaik, Kerr, & Statin, 2011). Moreover, exposure to a close reciprocal friend with internalizing problems has been found to reinforce adolescents’ tendencies toward depression (Giletta et al., 2011; Stevens & Prinstein, 2005), which has recently been reported in eight-year-olds also (Mercer & DeRosier, 2010). The latter requires replication in a sample of young children in order to ensure robustness of these findings. Thus, in order to increase understanding of the social risk factors implicated in the etiology of internalizing problems, it is important to establish whether internalizing problems cluster in friendships. To address this issue, the present study examines whether the degree of homophily in internalizing problems differs between reciprocated and unilateral friendships during early childhood.

Friendship and Internalizing Problems

Friends may serve as cognitive and social resources for children, for example by promoting self-esteem and general self-worth (e.g., Bagwell, Newcomb, & Bukowski, 1998; Hartup, 1996). Also, friends provide children with a context for social skills learning and serve as models for future relationship development (Bukowski, Newcomb, & Hartup, 1996). Most children have mutual friends (Hartup & Stevens,
1999), and children with internalizing problems are as likely to have reciprocal best friends as other children (Ladd & Burgess, 1999). Further, internalizing problems do not predict change in number of friends over time (Brendgen, Vitaro,Turgeon, Poulin, & Wanner, 2004; Ladd & Burgess, 1999). Contrary to these findings is that children with close relations may still suffer from internalizing problems, as internalizing problems tend to cluster in reciprocated friendships (Mercer & DeRosier, 2010). Indeed, negative peer experiences (e.g., peer rejection and victimization) are important predictors of internalizing problems; likewise, difficulties with friendships are associated with several adjustment problems, such as anxiety and depressed affect (Berndt, 1996; Bukowski et al., 1996; Hodges, Boivin, Vitaro, & Bukowski, 1999; Ladd & Kochenderfer, 1996; Rubin, Coplan, & Bowker, 2009).

It is widely thought that formation and maintenance of friendships are to a large part driven by preferences of similarities in appearances, such as age, gender, racial and ethnic background, behaviors, and opinions (i.e., selection processes; Aboud & Mendelson, 1996; Rubin, Lynch, Coplan, Rose-Krasnor, & Booth, 1994). This assumption is referred to in the literature as the ‘homophily hypothesis’ or ‘similarity attraction hypothesis’ (Berndt, 1982; Byrne, 1971; Kandel, 1978). It has been argued that reciprocated friendships may provide a primary context for mutual influence, because children have the greatest opportunity to interact and to share their personal feelings within these relationships (e.g., Rose, 2002). In accordance with this hypothesis, older children and adolescent friends have been found to resemble their reciprocal friends with regard to a host of externalizing behaviors (e.g., Engels, Vitaro, Den Exter Blokland, De Kemp, & Scholte, 2004; Poelen, Engels, Van der Vorst, Scholte, & Vermulst, 2007; Poulin et al., 1997; Prinstein, Boergers, & Spirito, 2001; Prinstein, Meade, & Cohen, 2003; Reitz, Dekovic, Meijer, & Engels, 2006; Vitaro, Tremblay, Kerr, Pagani, & Bukowski, 1997), shyness and prosocial behavior (Haselager, Hartup, Van Liershout, & Riksen-Walraven, 1998; Mrug, Hoza, & Bukowski, 2004), and to a lesser extent to internalizing problems such as depression (Brendgen et al., 2010; Hogue & Steinberg, 1995; Prinstein, 2007; Stevens & Prinstein, 2005) and anxiety (Rubin, Wojslawowicz, Rose-Krasnor, Booth-LaForce, & Burgess, 2006). In general, less attention has been directed to homophily with regard to internalizing problems in friendships and peer groups (see Giletta et al., 2011; Van Zalk et al., 2011), and specifically for young children, with one exception. Recently, children’s levels of social anxiety, loneliness, and depression have been found to become more similar to the average level of their friends in eight-year-olds (Mercer & DeRosier, 2010). Hitherto, thus only one study explicitly concentrated on friendship similarities in internalizing problems in children. However, while similarities regarding externalizing problems in friendships have been extensively reported and internalizing problems are highly co-morbid with externalizing problems in early childhood (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003), this study did not control for externalizing problems. The reported findings may have been attributable to concurrent externalizing problems. Therefore, we will control for externalizing problems in all analyses.

### Mechanisms of Homophily in Internalizing Problems during Early Childhood

Why would we expect homophilous friendships regarding internalizing problems during early childhood? Several mechanisms have been put forth explaining homophily of internalizing problems in adolescence, such as co-rumination, adoption of depressogenic attributional styles, negative affect induction, interpersonal stressor generation, and operation of network processes (Coyne, 1976; Rose, 2002; Rudolph, Flynn, Abaied, Groot, & Thompson, 2009; Schaefler, Kornienko, & Fox, 2011). More specifically, some of these mechanisms may underlie selection processes (i.e., children with problems actively seek out others to befriend with similar problems) and ‘others’ socialization’ (i.e., close friends reinforce each other’s problem behaviors). Due to the developmental level of 4-8-year-olds, not all of these mechanisms may apply. Mechanisms focusing on emotional contagion of internalizing problems may be most apt. Negative affect induction states that depressed individuals display maladaptive interactional styles that may induce depressed mood in others (Coyne, 1976), and relatedly, interpersonal stressor generation posits that depressed children generate stress in their relationships which may not only promote or exacerbate their symptoms but also affect those of their relational partners (Rudolph et al., 2009). Indeed, more conflict and less collaboration are displayed in interactions of depressed children with peers in comparison to those of non-depressed children (Rudolph et al., 2000). In addition, ‘age-specific manifestation’ of symptoms of internalized distress, such as irritability, sadness, guilt, anhedonia and somatic complaints (Luby, 2010), may produce negative reactions in peers (Coyne, 1976). In line with this reasoning is a recent study that showed depression contagion in 8-year-olds and adolescents. However, whereas co-rumination mediated depression contagion among adolescents, it did not in children (Schwartz-Mette & Rose, 2012). Co-rumination, which is thought to affect friends’ internalizing problems by extensively discussing and re-visiting problems, speculating about problems, and focusing on negative feelings (Rose, 2002), requires advanced cognitive capabilities such as perspective-taking and abstract thought, and may thus be not applicable to young children. Children below the ages of 7-8 have generally not been found to possess complex cognitive capabilities (Best & Miller, 2010), although a meta-analysis showed that 6-year-olds’ attributional style and depression are correlated (Joiner & Wagner, 1995). However, young children’s friendships are less intimate and characterized by less self-disclosure (McDougall & Hymel, 2007). Thus, chances are limited that children will engage in complex interactions that reveal their depressogenic attributional styles, and that friends will adopt these attributional styles (Stevens & Prinstein, 2005). All in all, we argue that the latter
cognitive mechanisms may not be as plausible as the former emotional mechanism underlying homophily of internalizing problems during early childhood.

Similarity in internalizing problems between children involved in a friendship may result from socialization processes, such as emotional contagion, but also from selection processes. That is, similarity in internalizing problems may affect children’s formation of new relationships. In this regard, children with internalizing problems have been found to have problematic interactions with peers, which is also associated with social withdrawal, and social withdrawal with internalizing problems, although directionality of effects is somewhat unclear (Rubin et al., 2009). Withdrawal is also associated with a reduced pool of friends to befriend, less chances of transivity (i.e., friendships with a friend of a friend), and fewer friendship ties through popularity. Thus, withdrawn children, who may also display internalizing problems, are likely to befriend children outside the normative network processes (Schaefer et al., 2011). Following this reasoning, this may lead to higher chances of friendship with other marginalized children, such as children who exhibit internalizing problems and less chances of friendship with children in the normative peer group. Thus, although children with internalizing problems may wish to befriend children who do not necessarily have internalizing problems, they may not succeed. Therefore, in a sort of ‘default selection process’, children with internalizing problems may be more likely to end up being friends with other children showing internalizing problems (Hektner, August, & Realmuto, 2000). As such, homophily is expected in reciprocated, but not unilateral friendships.

Study Objectives

The present study aimed to identify homogeneity in friendships concerning internalizing problems in 4-8 year old children from a large nation-wide Dutch sample. As research is very scarce regarding similarity in friendships of young children, this study is of an explorative nature. First, based on theory and empirical findings during adolescence, we expected to find support for the homophily hypothesis in friendships of young children, with stronger homophily expected in reciprocal friendships versus unilateral friendships (Giletta et al., 2011; Giletta et al., 2012). Thus, children who nominate a peer as a friend who does not return the nomination, i.e., unilateral friendships, are thought to be less similar to the nominee’s internalizing problems. On the contrary, two children who nominate each other as friends, i.e., reciprocal friendships, are thought to be more similar to each other in their internalizing problems. Second, as gender is one of the most powerful and universal dimensions for homophily, it is deemed important to assess the moderating role of gender (Mehta & Strough, 2009). During adolescence, gender differences are evident due to the increase in internalizing symptoms in females but not males, and females having more intimate friendships than males (Rose & Rudolph, 2006; Twenge & Nolen-Hoeksema, 2002). These differences in prevalence rates of internalizing problems and intimate friendships between boys and girls are less accentuated during early childhood, thus, we hypothesize that gender differences would be absent in homophily of internalizing problems. In addition to assessing the role of same-gender friendships, we also investigate the moderating role of mixed-gender friendships as this is very rarely studied (Bleske-Rechek & Buss, 2001), especially in childhood and regarding internalizing problems. Finally, we control for externalizing problems, as comorbidity rates are high in childhood (Costello et al., 2003).

Method

Participants

Participants were 1,584 Dutch children (M age = 5.9, SD = 1.26, 49.6 % boys) enrolled in the Kind in Zicht study. These children were selected from 2,558 children who were eligible for the study. Parental consent was obtained for 2,360 children. Subsequently, teachers administered questionnaires for all children in their class, resulting in a number of returned questionnaires of 2,182. Of the children with parental consent, 1,832 participated in sociometric interviews, 1,682 could be included in a dyadic friendship. Finally, for 49 dyads no teacher data were available leading to a final sample of 792 dyads or 1,584 children. The majority of teachers (99.5 %) and children (94.7 %) were of Dutch origin. Children were recruited for the study via 29 primary schools throughout the Netherlands. Four grades were included in the study; the preschool and kindergarten year, first and second grade. In total, 110 classrooms participated. Class sizes varied from 7 to 36. Parents were informed via letters which were handed out from the schools. Subsequently, passive informed consent was collected with 8.4 % of parents refusing to participate. Friendship nomination interviews were conducted from January until March 2010 outside of the classroom. A total of 1,832 children (83.9 %), ranging in age from 3.75 to 9.08 years (M = 6.04, SD = 1.28) at the time of nomination, had complete friendship nomination interview data. Of these children 50.3 % were boys.

Procedures

The Kind in Zicht study is a cohort study of Dutch children aged 4-8. During the school year prior to data collection, 2008-2009, schools were randomly selected from the population of elementary schools in the Netherlands. Schools in the larger provinces, Noord-Holland, Zuid-Holland, Noord-Brabant and Gelderland and the four largest cities, Amsterdam, Rotterdam, The Hague, and Utrecht, were oversampled. In total, 440 schools were selected. Principals of these schools received a letter inviting them for the study. Subsequently, principals were asked for participation by...
phone, which led to participation of 29 schools (6.6\%), containing 2,558 children in preschool and kindergarten classes, grade 1 and 2 classes. Only these grades were eligible for participation. Schools received 1000 Euros for their participation. During the school year of data collection, 2009-2010, teachers handed out information and consent letters to parents. Consent of 2,360 (92\%) parents was obtained. Teachers completed the questionnaires digital or by paper and pencil for all the children in their class, whose parents gave consent. There were no differences depending on mode of administration for the emotional problems $t(1908) = -.94, p = .35$; somatic complaints $t(1291) = 1.51, p = .13$, anxious-depressed $t(1290) = 1.18, p = .24$; withdrawn $t(1280) = 1.56, p = .12$, and externalizing scales $t(1216) = -.46, p = .64$. Sociometric interviews were held at schools from January until March 2010 in which 1,832 (77.6\%) children participated.

Children were considered to be a member of a reciprocal friendship dyad when the classmate they nominated as a friend reciprocated the nomination as friend. Children were considered unilateral friends when the classmate they nominated did not reciprocate the nomination. As children could name up to five friends, some participants were involved in more than one friendship at a time. There were 1,047 children (62.4\%) who named fewer than 5 friends. The inclusion of all friendship dyads in the analyses would have introduced the potential for bias due to unequal contributions by individuals. Thus, to avoid these problems, analyses focused on the highest ranked reciprocal friends and each participant was restricted to one and only one friendship dyad ($n = 1,682$. $M$ age = 5.85, $SD$ = 1.26, 49.6\% boys). The first criterion used to create dyads was the highest rank; the second was that we aimed to create as many dyads as possible, that is to match as many participants as possible in friends’ dyads to avoid excluding them from the analyses. Finally, when it was not possible to decide, we randomly chose in which dyads participants were included. Of these dyads, 49 did not have complete teacher reports leading to a final sample of 792 dyads. Chi-square tests were performed in order to examine whether children with a reciprocal or unilateral friendship differed from excluded children on socio-demographic characteristics (i.e., gender, age, ethnicity, educational level, family structure and marital status) and $t$-tests were used to test whether level of internalizing and externalizing problems differed between included and excluded children. Significant effects emerged for age $\chi^2(8, N = 2182) = 43.72, p < .00$, ethnicity $\chi^2(4, N = 1453) = 58.48, p < .00$, educational level $\chi^2(7, N = 1503) = 38.29, p < .00$, internalizing $t(2236) = 2.66, p = .008$, and externalizing problems $t(2236) = 3.57, p < .001$. Young children with the Dutch nationality, mothers with a high educational level and low levels of internalizing and externalizing symptoms were overrepresented among participants who had a reciprocal or unilateral friend versus children who did not have these friendships. No significant effects emerged for gender, family structure, and marital status.

**Measures**

**Strengths and difficulties questionnaire.** The Dutch teacher informant version of the SDQ was used to assess internalizing and externalizing behavior (SDQ; Widenfelt, Goedhart, Treffers, & Goodman, 2003). The subscales measuring emotional symptoms (e.g. many worries, often seems worried) and conduct problems (e.g. often lies or cheats) each contain five items. Teachers rated children on a 3-point scale ranging from 0 (not true) to 2 (certainly true). The scoring procedures used in this study are available online (Widenfelt et al., 2003). Cronbach’s alpha was $\alpha = .71$ for emotional symptoms ($M = .97, SD = 1.54$) and $\alpha = .53$ for conduct problems ($M = .76, SD = 1.23$). Reliabilities based on Structural Equation Modeling -known as Jöreskog rho or McDonalds Omega (Jöreskog, 1971; McDonald, 1978.1999; Revelle & Zinbarg, 2009; Stone et al., in press) were $\alpha_1$ and $\alpha_2$, respectively. To control for conduct problems, the emotional problems scale was regressed on the conduct problems scale, creating a residual emotional problems score. The distribution of the residual score was skewed, and therefore scores above the 99th percentile were rescaled to the 99th percentile value.

**Teacher report form.** The Dutch versions of the C-TRF and TRF were used to assess internalizing and externalizing behaviour as reported by teachers (Achenbach & Rescorla, 2000; Achenbach & Rescorla, 2001; Verhulst, Van der Ende, & Koot, 1997). The C-TRF, used for children aged 1.5-5 years, comprises 100 items; the TRF targets 5-18-year-olds and consists of 118 items. These items are rated using a 3-point Likert scale, where 0 indicates responses of ‘“not true”’, 1 ‘“somewhat or sometimes true”’, and 2 ‘“very true or often true”’. The C-TRF and TRF somatic complaints, anxious/depressed and withdrawn subscales were aggregated to enable comparability with the SDQ. The emotionally reactive subscale from the C-TRF was removed from the analyses as this scale was not present in the TRF version. The three subscales comprising the internalizing scale (i.e., somatic complaints $\alpha = .51$, $SD = .87$), anxious/depressed and withdrawn $\alpha = .76$, $SD = .97$), and withdrawn $\alpha = .74$, $SD = .77$, and the externalizing scale $\alpha = .93$, $SD = .83$, $SD = .64$) respectively for C-TRF and TRF were used in the current study. Again, to control for externalizing problems, the somatic complaints, anxious/ depressed and withdrawn scales were, separately, regressed on the externalizing problems scale, thus creating three residual scores. The distributions of the residual scores were skewed, and therefore scores above the 99th percentile were rescaled to the 99th percentile value.

**Friendship nomination.** A sociometric one-on-one interview was used to identify friendship dyads. A trained research assistant picked the children up from their classroom and took them to a quiet place in the school to sit down. The child was asked to nominate five of his / her best friends from class: “Who are your best friends in class? You can name five.”. After naming classmates, children were asked to order...
these friends: “Who of these friends you just named is your very best friend?” This question was repeated until the five named friends were ordered. The child was allowed to name fewer than five friends. The child then returned to the classroom and took the next child to the research assistant. Responses of the child were written down verbatim by the interviewer.

Ethnicity. Parents were asked to indicate what the child’s nationality was, with the question “What is your child’s nationality?”. Answering options were ‘Dutch’, ‘Moroccan’, ‘Turkish’, ‘Surinam’ or ‘different’.

Strategy for analysis
First, correlations were computed between the TRF Somatic Complaints, Anxious-Depressed, Withdrawn and Externalizing Problems and the SDQ Emotional Problems scales. The unmodified scales were used when we computed these correlations. Subsequently, intraclass correlations between friends’ scores (both reciprocal and unilateral) were computed with the residual scores of each scale. To deal with the random assignment of some children as targets and others as friends (i.e., the indistinguishable nature of the dyad partners), a pairwise approach recommended by Griffin and Gonzalez (1995) was used. This requires all relationships to be entered twice, once with the target child’s scores entered first and the friend’s score entered second, and once with the friend’s scores entered first and the target child’s scores entered second. Therefore, intraclass correlations were computed to assess the degree of similarity between dyad members. In order to account for interdependency between dyad members, the statistical significance of the pairwise correlations was adjusted according to the procedure proposed by Griffin and Gonzalez (1995) (see pag. 432).

Subsequently, a Confirmatory Factor Analysis (CFA) was performed in MPlus with the internalizing subscales as indicators. Residual scores of the TRF Somatic Complaints, Anxious-Depressed, Withdrawn, and the SDQ Emotional Problems scales were used as input for the CFA. CFA enables a) obtaining one latent indicator of internalizing problems and b) removal of measurement error leading to a more pure internalizing factor (Bentler & Kano, 1990; Crocker & Algina, 1986). For these analyses, we used maximum likelihood estimation with the Huber-White covariance adjustment (MLR in Mplus 5.0: Muthén & Muthén, 1998-2007). Model fit was assessed with various fit indices, including robust chi-square with estimated degrees of freedom (df), comparative fit index (CFI; Bentler, 1990), root mean squared error of approximation (RMSEA; Byrne, 1998), and Tucker-Lewis index (TLI; Tucker, & Lewis, 1973).

Second, the latent internalizing factor scores were saved and used in subsequent dyadic analyses. Correlations between friendship types (reciprocal, unilateral) were computed for the latent internalizing factor and the four indicators. In order to test gender moderation, correlations were computed separately for these variables, while keeping the distinction between reciprocal and unilateral friends. A series of analyses involving Fisher’s r to z transformations were conducted to test the differences between correlations. Differences were tested for a) friendship types b) gender between friendship types and c) gender within friendship types.

Results
Descriptive statistics
Correlations and intraclass correlations are presented for 792 dyads (557 reciprocal, 235 unilateral) in Table 1. In order to control for Externalizing Problems, partial intraclass correlations of Emotional Problems, Somatic Complaints, Anxious-Depressed and Withdrawn were computed and reported in Table 1. All study variables were positively related, with the internalizing subscales emotional problems, somatic complaints, anxious-depressed and withdrawn variables generally showing higher correlations with each other, than with the externalizing scale. The intraclass correlations at the subscale level show that, across all dyads, there is some evidence for similarity, specifically for the subscales somatic complaints, anxious-depressed, withdrawn, and externalizing. The partial intraclass correlations show that similarity for the three former subscales remains significant after controlling for externalizing problems.

Latent Dyadic Factor Model Results
Confirmatory factor analysis with four indicators, emotional problems, somatic complaints, anxious-depressed, and withdrawn loading on a latent internalizing factor yielded acceptable fit ($\chi^2 (2) = 2.60, p = 0.27$; $\text{CFI} = .977$; $\text{RMSEA} = .034$ (CI = .000-.133); $\text{TLI} = .921$). To assess the degree of homophily within dyads involving reciprocated and unilateral friendships, intraclass correlations were computed for the latent internalizing factor scores, emotional problems, somatic complaints, anxious-depressed, and withdrawn scores. The intraclass correlations were computed separately within group, and also separately by gender, and are shown in Table 2.

For reciprocal friends, significant correlations were found, and Fisher’s r to z analyses revealed that these differed significantly from correlations found for unilateral friends, respectively. For the latent internalizing factor ($z = 2.23, p < .05$), somatic complaints ($z = 2.52, p < .05$), and anxious-depressed ($z = 2.44, p < .05$), but not for the emotional problems ($z = 1.94, p > .05$) and withdrawn scale ($z = -2.1, p > .05$). Correlations for unilateral friendships were all non-significant. Thus, children within a reciprocated friendship showed greater behavioral similarity than children within a unilateral friendship.
moderated friendship types

In order to examine whether differences between reciprocal and unilateral friendships also differed for boys and girls, correlations were computed separately for gender. For boys’ reciprocal friendships, significant correlations were found, and these differed significantly from correlations found for unilateral friends, respectively, for the latent internalizing factor (z = 2.54, p < .05) and anxious-depressed (z = 2.45, p < .05), but not for the emotional problems (z = .44, p > .05), somatic complaints (z = 1.33, p > .05) and withdrawn scales (z = .76, p > .05). For girls’ reciprocated friendships, significant correlations were found for two subscales, but these did not differ significantly from correlations for girls’ unilateral friendships, for the somatic complaints (z = 1.60, p > .05), and anxious-depressed (z = 1.01, p > .05) scales. Further, correlations for both boys’ and girls’ unilateral friendships were all nonsignificant. For mixed-gender reciprocated friendships, a significant correlation was found only for the withdrawn scale and, remarkably, for unilateral mixed-gender friendships the correlation for the withdrawn scale was also significant. The correlations for reciprocated and unilateral friends did not differ significantly (z = .95, p > .05), indicating that children in reciprocal mixed-gender friendships show equally great behavioral similarity as children in unilateral mixed-gender friendships.

Moderation within friendship types

Analyses conducted separately by gender show that both boys’ and girls’ reciprocated friendships are characterized by similarity in the somatic complaints and anxious-depressed scales. Further, reciprocal male friendships are similar in the latent internalizing and emotional problems scales. None of the correlations differed between boys’ and girls’ friendships, nor for reciprocal or unilateral friendships. These findings indicate that boys and girls do not differ within friendship type in their similarity. The mixed-gender reciprocated friendships are characterized by similarity in the withdrawn scale. The correlations did not differ between mixed-gender and boys’; and girls’ reciprocal or unilateral friendships, except for the correlations between unilateral mixed-gender and unilateral male friendships (z = 2.86, p < .01). These findings indicate that mixed-gender gender friendships do not seem to differ within friendship type in their similarity, although they do differ regarding withdrawn behavior. Unilateral mixed-gender dyads are more similar to each other regarding withdrawn behavior than unilateral male dyads.

Discussion

Homophily of normative and maladaptive developmental characteristics has been very rarely studied in early or middle childhood (Dunn & Cutting, 1999; Eivers, Brendgen, Vitaro, & Borge, in press; Gleason, Gower, Hohmann, & Gleason, 2005; Haselager et al., 1998), let alone homophily of internalizing problems (Mercer & DeRosier, 2010). Furthermore, no studies yet have tested homophily in internalizing problems and controlled for externalizing problems while a) comorbidity rates are

---

**Table 1** Correlations between Indicators of the Latent Internalizing Factor (N = 1,584) and Intraclass Correlations (N = 792)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Intraclass r</th>
<th>Intraclass-partial r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Emotional problems</td>
<td>-</td>
<td>.07</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Somatic complaints</td>
<td>.44**</td>
<td>-</td>
<td>.20**</td>
<td>.15**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Anxious-depressed</td>
<td>.72**</td>
<td>.37**</td>
<td>-</td>
<td>.17**</td>
<td>.15**</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Withdrawn</td>
<td>.38**</td>
<td>.26**</td>
<td>.39**</td>
<td>-</td>
<td>.17**</td>
<td>.10*</td>
</tr>
<tr>
<td>5</td>
<td>Externalizing</td>
<td>.21**</td>
<td>.16**</td>
<td>.21**</td>
<td>.38**</td>
<td>-.33**</td>
<td>-</td>
</tr>
</tbody>
</table>

**p < .01, * p < .05

**Table 2** Correlations between Indicators of the Latent Internalizing Factor (N = 1,584) and Intraclass Correlations (N = 792)

<table>
<thead>
<tr>
<th>Friend dyads</th>
<th>Total</th>
<th>Boys</th>
<th>Girls</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocal (n = 557, 242 boys, 64 mixed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent internalizing</td>
<td>.15**</td>
<td>.19**</td>
<td>.01</td>
<td>.23</td>
</tr>
<tr>
<td>Emotional problems</td>
<td>.11**</td>
<td>.16**</td>
<td>.05</td>
<td>.15</td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>.20**</td>
<td>.24**</td>
<td>.23**</td>
<td>.00</td>
</tr>
<tr>
<td>Anxious-depressed</td>
<td>.21**</td>
<td>.23**</td>
<td>18**</td>
<td>.22</td>
</tr>
<tr>
<td>Withdrawn</td>
<td>.09*</td>
<td>.01</td>
<td>.12</td>
<td>.27*</td>
</tr>
<tr>
<td>Unilateral (n = 235, 96 boys, 45 mixed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latent internalizing</td>
<td>-</td>
<td>-12</td>
<td>.03</td>
<td>-03</td>
</tr>
<tr>
<td>Emotional problems</td>
<td>-04</td>
<td>-11</td>
<td>-02</td>
<td>-05</td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>.01</td>
<td>.08</td>
<td>.03</td>
<td>-16</td>
</tr>
<tr>
<td>Anxious-depressed</td>
<td>.02</td>
<td>.07</td>
<td>.06</td>
<td>.01</td>
</tr>
<tr>
<td>Withdrawn</td>
<td>.11</td>
<td>.08</td>
<td>.16</td>
<td>.41**</td>
</tr>
</tbody>
</table>

**p < .01, * p < .05
high in childhood (Costello et al., 2003) and b) similarities in externalizing problems in friendships of children have been reported (e.g., Dishion, Patterson, & Griesler, 1994). The current study extends prior research by exploring similarity in internalizing problems between young children and their friends, with the goal of increasing understanding of social risk factors implicated in the etiology of internalizing problems. Our results support our hypotheses, in that similarity was present in reciprocal, but not unilateral, friendships of young children. These results are in accordance with the adolescent literature, where evidence showing that homophily in internalizing problems is present, is accumulating (Brendgen et al., 2010; Gilletta et al., 2011; Hogue & Steinberg, 1995; Prinstein, 2007; Stevens & Prinstein, 2005; Van Zalk et al., 2011), and with studies focusing on children (Mercer & DeRosier, 2010; Rubin et al., 2006). However, in these studies, children were substantially older than in the current study (i.e., 8-10 years of age) and the sample size in the Rubin et al. study was quite small. To our best knowledge, this study is the first to show that there is a link between internalizing problems and friendships of 4-8-year-olds. Although the correlations reported in the current study are not as high as those found in studies with older children or adolescents (Brendgen et al., 2010; Hogue & Steinberg, 1995; Prinstein, 2007; Rubin et al., 2006; Stevens & Prinstein, 2005), we think the magnitude of these correlations may be explained by the age of the children. As internalizing problems tend to peak during adolescence (e.g., Cole et al., 2002; Keenan-Miller et al., 2007; Twenge & Nolen-Hoeksema, 2002), more variance in these problems can be explained. In our opinion the limited variance found in early childhood internalizing problems may account for the small correlations reported in this study.

Second, as internalizing problems often co-occur with externalizing problems, it was deemed important to establish whether internalizing problems cluster in friendships of young children, as externalizing problems do (Angold, Costello, & Erkani, 1999; Dishion et al., 1994). Because we controlled for externalizing problems, our results support homophily of purely internalizing symptoms among friends. Prior studies on homophily of internalizing problems did not take concurrent externalizing problems into account (Mercer & DeRosier, 2010; Rubin et al., 2006), and reported effects might have been, at least partially, attributable to externalizing problems.

Further, our results show that homophily in friendships is present regardless of gender, indicating the robustness of the reported findings. Regarding mixed-gender friendships, there was some evidence for homophily. To our best knowledge, homophily in internalizing symptoms within mixed-gender dyads of children has not been specifically examined yet. Also, we found that these children resembled each other most strongly in withdrawn behavior, both in reciprocated and unilateral friendships. This suggests that children in mixed-gender friendships (i.e., reciprocal) and children who want to be friends with a child of the opposite gender (i.e., unilateral) are both reported to show withdrawn behaviors. Although speculative in nature, extant literature may explain these findings. Most children consider same-gender friendships and play styles more acceptable than being friends with children of the other gender or having a play style of the other gender (Mehta & Strough, 2009). Having a mixed-gender friendship could be associated with gender atypical behavior, and evidence supports the notion that children react negatively to atypical gender behavior of other children (e.g., Rubel et al., 2007). This may be associated with withdrawn behavior. Finally, as a substantial part (13.7%) of all friendships were mixed-gender in the current study, future research on mixed-gender friendships is warranted.

Although we were not able to test the mechanisms explaining homophily (i.e., selection and socialization processes), the current findings provide some indirect insights into the nature of homophily in friendships of young children. Even young children’s reciprocal friendships are characterized by similarity in internalizing problems, whereas their unilateral friendships are not. Thus, friendship may not always serve a protective function (cf. Rubin et al., 2006), and possibly, for some children, it may be conceptualized as a risk factor for early internalizing problems. Indeed, although socially withdrawn children have been found to be as likely as non-withdrawn children to have mutual friends, the quality of their friendships has been reported to be lower (Rubin et al., 2006). Future studies should address whether all children in homophilous internalizing friendships continue to develop symptoms of depression and anxiety, or whether some do not, and thus to what extent homophilous internalizing problems pose a risk for subsequent development (e.g., shy and socially withdrawn children; Gazelle & Rudolph, 2004). Further, from these findings, we might infer that processes other than co-rumination might underlie homophily, as it is unlikely that children this age are cognitively able to co-ruminate (Best & Miller, 2010). Although it is not evident from our findings that these children may actually socialize each other into higher levels of internalizing symptoms, the adolescent literature points to consistent evidence for socialization effects (e.g., Gilletta et al., 2011) as has been reported in 8-10-year-olds also (Mercer & DeRosier, 2010). Children with internalizing problems may induce depressed mood in their friends, and vice versa, and generate stress in their relationship, which may increase both the children’s symptoms (Coyne, 1976; Rudolph et al., 2009).

As for selection processes, whether children with internalizing problems actively seek out others to befriend with similar problems, or whether they unintentionally end up with those children is not clear yet. The similarity attraction hypothesis explicitly, and the homophily hypothesis more implicitly, assume that children befriended based on preferences for similarity (Berndt, 1982; Byrne, 1971; Kandel, 1978). In contrast, the default selection hypothesis states that friendship selection does not occur on the basis of similarity, but is due to a lack of availability instead of actual preference for similarity (Hektner et al., 2000). Although ample studies support the homophily
CHAPTER 6 FRIENDSHIP SIMILARITIES IN INTERNALIZING PROBLEMS

The hypothesis regarding similarity of behavioral characteristics (e.g., Giletta et al., 2011; Mercer & DeRosier, 2010; Schaefer et al., 2011), support for its preference thesis is limited. Research on the default selection hypothesis is scarce, but two recent studies support the tenet that children prefer others as friends who are not similar to them in terms of psychopathology, but still are friends with those children that do show similar psychopathology (Schaefer et al., 2011; Sijtsma, Lindenberg, & Veenstra, 2010). By testing these two competing hypotheses against each other (cf. Sijtsma et al., 2010), the causes for formation of homophilous friendships regarding internalizing problems could be elucidated.

Although this study has important features, some limitations should be noted. First, this study has a cross-sectional design, which did not allow us to test possible mechanisms, such as selection and socialization. A longitudinal design would allow us to examine who were not friends at Time 1 but become friends at Time 2 (i.e., selection) and to examine if friends become more similar to each other over time (i.e., socialization) (Giletta et al., 2011). Further, future longitudinal studies are needed to assess when friends become similar to each other in internalizing problems, as is shown in adolescence (Brendgen et al., 2010; Giletta et al., 2011; Hogue & Steinberg, 1995; Prinstein, 2007; Stevens & Prinstein, 2005; Van Zalk et al., 2011). Second, our study was conducted in a general population where rates of internalizing problems are relatively low. Replication of this study in a clinical sample is important to a) explore whether clinically depressed children at this age have friends who are similar to themselves and if so, b) to investigate why children befriend other depressed children.

Third, this study used teacher reports to assess problem behaviors. The low concordance between teacher and self-reports regarding internalizing problems is well-established (Achenbach, McConaughy, & Howell, 1987), so self-reports may be used in future studies to investigate the homophily principle as well. Reliable and age-appropriate instruments that use self-reports are available and can be used for this purpose, for children from age 4, such as the Berkeley Puppet Interview, and the Dominic Interactive for school-aged children (BPI; Measelle, Ablow, Cowan, & Cowan, 1998; DI; Valla, 2000). It is recommended to use parent reports to complement teacher- and self-reports, as a multi-informant approach seems to be most reliable in assessing internalizing problems (De Los Reyes & Kazdin, 2005; Stanger & Lewis, 1993). Fourth, reports of symptoms of internalizing problems via questionnaires may be too distal for measuring homophily in young children. Observations of interactions between friends and what these interactions characterize may be more suited, as non-verbal and verbal behavior encompasses information related to cognitions underlying internalizing problems. Relatedly, the reliability of the scale somatic complaints was low, therefore results for this subscale specifically should be interpreted with caution. Finally, the dyads were selective in terms of age, nationality, educational level, and internalizing and externalizing symptoms. These findings may be due to the relatively low participation rate of schools in our study. In the Netherlands, these participation rates are quite typical, but with the use of more incentives for the school, participation rates may become higher. Future research should attempt to include children with different nationalities, low SES backgrounds, and higher levels of internalizing and externalizing problems in order to generalize the findings to these populations.

Despite these limitations, results from this study may have potential implications for practice. Although it is not clear from these findings whether children actually socialize each other into higher levels of internalizing problems, the finding that internalizing problems cluster in friendships of these young children marks the importance of early intervention. Pairing children with psychopathological symptoms with those without symptoms seems a fruitful approach. Results from a summer treatment program targeting aggressive children are positive, such that children’s aggression decreased when paired to a non-aggressive counterpart, while the aggressive child did not evoke aggression in the non-aggressive child (Hektner, August, & Realmuto, 2003). Of course, these interventions targeting aggressive children need to be tailored to the specific needs of children exhibiting internalizing problems. Specifically, some techniques which may be effective for children with externalizing symptoms, may not work for children with internalizing symptoms (Granic, 2012). Treatment manuals focusing on internalizing problems mention the importance for children to get habituated to anxiety adequately before terminating exposure of fear-inducing stimuli (Kendall & Hedtke, 2006). Therefore, treatments targeting children with internalizing problems should utilize a fear-hierarchy, whereby the child is only exposed to an interaction with the paired child if this feels safe for the child with internalizing symptoms (Chorpita, 2007). In the Hektner et al. study, paired children in the summer treatment program played a competitive game together. This competitive game in itself may be fear-inducing for children with internalizing symptoms. Depending on the specific symptoms of the target child, an interaction should be tailored such that this is not fear inducing.
References


Chapter 7

The Co-Development of Parenting Stress and Childhood Internalizing and Externalizing Problems

Submitted for publication:
Stone, L. L., Mares, S., Otten, R., Janssens, J. M. A. M., & Engels, R. C. M. E.
CHAPTER 7

Abstract

Although the detrimental influence of parenting stress on problem behaviors is well established, it remains unknown how these constructs affect each other over time. Therefore, from a transactional model, this study investigates how the development of internalizing and externalizing problems is related to the development of parenting stress in children aged 4-9. Mothers of 1,582 children aged 5.08 (SD = 1.25) participated in three one-year interval data waves. Internalizing and externalizing problems, and parenting stress were assessed by maternal self-report. Interrelated development of parenting with internalizing and externalizing problems was examined using Latent Growth Models. Directionality of effects was further investigated by using cross-lagged models. Parenting stress and externalizing problems showed a decrease over time, whereas internalizing problems remained stable. Initial levels of parenting stress were related to initial levels of both internalizing and externalizing problems. Decreases in parenting stress were related to larger decreases in externalizing problems and to the (stable) course of internalizing problems. Some evidence for reciprocity was found such that externalizing problem were associated with parenting stress and vice versa over time, specifically so for boys. Our findings support the transactional model in explaining psychopathology.

Introduction

Internalizing and externalizing problems that occur in adolescence and persist throughout adulthood are often rooted in childhood (internalizing: Mazza et al., 2009; Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; externalizing: e.g., Loeber & Hay, 1997; Ashford, Van Lier, Timmermans, Cuijpers, & Koot, 2008; Maggs, Patrick, & Feinstein, 2008). Specifically, while internalizing problems in childhood have been linked to pervasive and adverse developmental outcomes, such as depression and anxiety disorders, academic underachievement, and problems with employment (Aronen & Soininen, 2000; Woodward & Fergusson, 2001), externalizing problems in childhood increase the risk for aggression and substance use later in life (e.g., Loeber & Hay, 1997; Maggs, Patrick, & Feinstein, 2008). Therefore, understanding which early childhood factors are implicated in the development of in- and externalizing problems is paramount.

Whether in- or externalizing problems in childhood occur or not depends on a variety of individual and environmental factors and the interplay between them. As an environmental factor, parenting can be considered as the most important early childhood factor. Parenting a child may be a significant source of stress and according to Belsky’s ecological framework, contextual sources of stress can affect parenting negatively, in turn increasing child psychopathology (Belsky, 1984). Parenting stress is one of the most prominent sources of stress, as all parents experience parenting stress to some degree (Crnic & Greenberg, 1990; Hakvoort, Bos, Van Balen, & Hermanns, 2012) and as such it is an important avenue for research. Current theories of parenting stress either conceptualize stress as a disorder, which is in accordance with a categorical view (i.e., yes or no; stress as the consequence of stressful life events). In contrast, Daily Hassles Theory (Crnic & Greenberg, 1990) posits that stress is a typical process, wherein stress ranges on a continuum from mild everyday stressors to very severe stress (Deater-Deckard, 2004). The latter perspective is taken in the current study.

Studies have shown that there is indeed a positive link between parenting stress and internalizing problems, with these studies having a cross-sectional design (Anthony, Anthony, Glanville, Nairn, Waanders, & Shaffer, 2005; Costa, Weems, Pellerin, & Dalton, 2006; Hart & Kelley, 2006; Mesman & Koot, 2000; Rodriguez, 2011). Increasingly, longitudinal studies confirm these findings (Ashford, Smit, Van Lier, Cuijpers, & Koot, 2008; Bayer, Sanson, & Hempfl, 2006; Bayer, Hiscock, Ukoumunne, Price, & Wake, 2008; Mantymaa, Puura, Luoma, Latva, Salmelin, & Tamminen, 2012). However, these studies have all been conducted in the developmental period of preschool, while the role of parenting stress and its link to internalizing problems during early childhood is less well studied.

Regarding externalizing problems, the positive relation to parenting stress is well established cross-sectionally (Barry, Dunlap, Cotten, Lochman, & Wells, 2005; Blader,
may be expected that children whose parents show higher levels of parenting stress is positively related to internalizing and externalizing problems. It is hypothesized that parenting stress predicts externalizing and internalizing problems. As pressures from below have been shown to predict parenting behaviors other than parenting stress (Pomerantz & Eaton, 2001), we hypothesize that child psychopathology will also predict parenting stress. Fourth, boys are expected to display more externalizing problems than girls (Dishion & Patterson, 2006), while no gender differences are expected regarding internalizing problems (Ford, Goodman, & Meltzer, 2003).

Method

Sample and Procedure

Mothers of children aged 4-7 from 29 primary schools throughout the Netherlands were recruited for the Dutch “Kind in Zicht” study, in which mothers of 1,339 children participated (M age = 5.08, SD = 1.25, 50.1% boys) in the first assessment. At the subsequent assessment 979 mothers participated (67%), including 95 ‘new’ parents who did not participate in the baseline assessment. In the third assessment wave, 819 (61%) parents participated in the study, including 148 ‘new’ parents who did not participate in the baseline assessment. These ‘new’ parents were recruited for the study prior to the baseline assessment and gave consent to participate in the study, but simply did not participate in the baseline assessment. This is the reason why these parents are indicated here as being ‘new’. Due to use of Structural Equation Modeling, wherein missing cases are estimated, our final N is 1,582. At baseline, mothers had a mean age of 36.61 (SD = 4.41), the majority was of Dutch origin (92.4%) and were part of a two-parent household (89.1%). Most mothers, 44.6%, were highly educated with a college or university degree, 37.8% finished vocational education. 13.7% finished a low level of Dutch secondary school, and 4% finished a different form of education. Attrition analyses showed that families who completed three waves (n=817) did not differ from the dropouts (n=522) in child age, gender, maternal educational level, family structure, internalizing, and externalizing problems, and parenting stress. Families that completed three waves differed from the dropouts regarding ethnicity (OR 1.29, 95% CI 1.09 – 1.52, p = .003) with the dropouts being of greater non-Dutch origin.

We used data of three annual waves of Kind in Zicht, a large cohort study of Dutch children aged 4-7 at T1. Schools were randomly selected from the population of elementary schools in the Netherlands. Schools in the larger provinces, Noord-Holland, Zuid-Holland, Noord-Brabant and Gelderland and the four largest cities, Amsterdam, Rotterdam, The Hague and Utrecht, were oversampled. In total, 440 schools were selected. Principals of these schools first received a letter inviting expected that parents whose children show higher levels of problem behaviors decrease less in their levels of parenting stress. Third, regarding directionality, we hypothesize that parenting stress predicts externalizing and internalizing problems. As pressures from below have been shown to predict parenting behaviors other than parenting stress (Pomerantz & Eaton, 2001), we hypothesize that child psychopathology will also predict parenting stress. Fourth, boys are expected to display more externalizing problems than girls (Dishion & Patterson, 2006), while no gender differences are expected regarding internalizing problems (Ford, Goodman, & Meltzer, 2003).

Method

Sample and Procedure

Mothers of children aged 4-7 from 29 primary schools throughout the Netherlands were recruited for the Dutch “Kind in Zicht” study, in which mothers of 1,339 children participated (M age = 5.08, SD = 1.25, 50.1% boys) in the first assessment. At the subsequent assessment 979 mothers participated (67%), including 95 ‘new’ parents who did not participate in the baseline assessment. In the third assessment wave, 819 (61%) parents participated in the study, including 148 ‘new’ parents who did not participate in the baseline assessment. These ‘new’ parents were recruited for the study prior to the baseline assessment and gave consent to participate in the study, but simply did not participate in the baseline assessment. This is the reason why these parents are indicated here as being ‘new’. Due to use of Structural Equation Modeling, wherein missing cases are estimated, our final N is 1,582. At baseline, mothers had a mean age of 36.61 (SD = 4.41), the majority was of Dutch origin (92.4%) and were part of a two-parent household (89.1%). Most mothers, 44.6%, were highly educated with a college or university degree, 37.8% finished vocational education, 13.7% finished a low level of Dutch secondary school, and 4% finished a different form of education. Attrition analyses showed that families who completed three waves (n=817) did not differ from the dropouts (n=522) in child age, gender, maternal educational level, family structure, internalizing, and externalizing problems, and parenting stress. Families that completed three waves differed from the dropouts regarding ethnicity (OR 1.29, 95% CI 1.09 – 1.52, p = .003) with the dropouts being of greater non-Dutch origin.

We used data of three annual waves of Kind in Zicht, a large cohort study of Dutch children aged 4-7 at T1. Schools were randomly selected from the population of elementary schools in the Netherlands. Schools in the larger provinces, Noord-Holland, Zuid-Holland, Noord-Brabant and Gelderland and the four largest cities, Amsterdam, Rotterdam, The Hague and Utrecht, were oversampled. In total, 440 schools were selected. Principals of these schools first received a letter inviting
them to participate in the study and subsequently, were asked for participation by phone, which led to participation of 29 schools (6.6%), containing 2,558 children in two kindergarten classes, Grade 1 and 2. Schools received 1,000 for their participation. Teachers handed out information and consent letters to parents. Passive consent of 2,360 (92.3%) parents was obtained. Only mothers were allowed to participate in the study, as a mother is the primary caregiver in most families (Renk et al., 2003). In all waves, mothers completed questionnaires either digitally or by paper and pencil.

Measures

Internalizing and Externalizing Problems. The Dutch parent version of the Strengths and Difficulties Questionnaire was used at all waves to assess internalizing and externalizing problems (Van Widenfelt, Goedhart, Treffers, & Goodman, 2003). The subscale emotional symptoms (e.g., many worries, often seems worried) was used to measure internalizing problems. The conduct problems scale (e.g., often lies or cheats) was used to measure externalizing problems. Each scale contains five items and parents rated their child’s behavior on a 3-point scale ranging from 0 (not true) to 2 (certainly true). The scoring procedures used in this study are available online at www.sdqinfo.com: Higher scores indicate more problem behaviors. Cronbach’s alphas were .63, .67, .65 for emotional symptoms at T1, T2, and T3 and .48, .47, .55 for conduct problems at T1, T2, and T3. Alternative indicators of reliability based on Structural Equation Modeling – are known as Jöreskog rho or McDonalds Omega. Because this indicator is suggested to be more accurate when scale distributions are skewed, as is often the case in instruments measuring problem behavior, like the SDQ (Jöreskog, 1971; McDonald, 1978,1999; Revelle & Zinbarg, 2009; Stone, Otten, Ringlever, Hiemstra, Engels, Vermulst & Janssens, 2013) reliability was also calculated using omega (ω). Omega values were .79, .80, .81 at T1, T2, and T3 for the emotional symptoms scale, and .71, .75, .77 at T1, T2, and T3 for the conduct problems scale. Emotional symptoms and conduct problems were positively skewed and leptokurtic. In line with Tukey’s (1977) recommendations, the least strong transformation that yielded the most symmetric distribution was chosen for each scale. For the internalizing scale the square root was taken. A logarithmic transformation that yielded the most symmetric distribution was chosen for each scale. In line with Tukey’s (1977) recommendations, the least strong transformation that yielded the most symmetric distribution was chosen for each scale. For the internalizing scale the square root was taken. A logarithmic transformation that yielded the most symmetric distribution was chosen for each scale.

Parenting Stress. At all waves mothers rated the frequency of daily hassles with their child over the past 6 months (Parenting Daily Hassles: PDH; Crnic & Booth, 1991; Van der Wal. Van Eijden, & Bonsel, 2007). The questionnaire consists of 20 events of which the parent has to rate how often they occur (seldom, sometimes, often, constantly). A mean score was calculated with higher scores indicating higher parenting stress. Psychometric properties of the PDH have been found adequate (Crnic & Booth, 1991; Rispens, Hermanns, & Meeus, 1996). Cronbach’s alphas were .77, .78, .78 at T1, T2, and T3.

Mental Health. Finally, we control for maternal mental health, as this is strongly related to parenting stress (Patterson, 1982). The degree of mental health of the mothers during the past 4 weeks was measured at the first wave with a short version of the General Health Questionnaire (GHQ; Hardy, Shapiro, Haynes & Rick, 1999). Mothers rated their mental health via 12 questions (e.g., did you lose confidence in yourself? did you feel able to make decisions?) on a 4-point scale. A mean score was calculated, with higher scores indicating diminished mental health. Research into reliability and validity indicates that the GHQ has adequate psychometric properties (Koeter & Ormel, 1991). Cronbach’s alpha was .88.

Strategy for Analysis

First, means, standard deviations and bivariate correlations of all study variables were calculated. We investigated the growth of parenting stress, and internalizing and externalizing problems by employing univariate Latent Growth Curve modeling. Gender differences were investigated by comparing a freely estimated model to a model wherein parameters were constrained to be equal for boys and girls. If a significantly worse fit to the data was found for the constrained model, we employed a stepwise approach, such that each of the parameters were tested separately for gender differences. Subsequently, we tested whether the growth of parenting stress and internalizing problems and externalizing problems, respectively, were related to each other. The univariate growth models of internalizing and externalizing problems were combined with the growth model of parenting stress. Gender differences were tested by comparing a freely estimated model to a model wherein interrelations between internalizing/externalizing problems and parenting stress were constrained to be equal for boys and girls. Finally, to evaluate direction of effects of the associations of parenting stress with internalizing problems on the one hand, and externalizing problems on the other over time, we tested two cross-lagged path models, using MPLUS version 5 (Muthén & Muthén, 1998-2007). Again, gender differences were investigated by employing multi-group modeling. Cross-lagged paths and concurrent relations were constrained to be equal across boys and girls. A model wherein all parameters were estimated freely was compared to this constrained model. To test whether the constraints significantly deteriorated the model, a χ² difference test was used according to the recommendations on the website of Mplus in all models (see http://www.statmodel.com/chidiff.html). All models were controlled for maternal mental health and age. Because MPLUS uses a Full Information Maximum Likelihood principle, the analyses are based on all data points, which are 1,339 participants at T1 + 95 participants at T2 who did not complete the baseline assessment, +148 participants at T3 who did not complete the baseline assessment, leading to a final sample of 1,582 participants.
Since children from the same classes may share common behaviors (i.e., clustering), intraclass correlations (ICC) were calculated to determine the effects of class clustering. The ICC’s for internalizing problems were .019, .00, and .003 at T1, T2, and T3, respectively. .00, .047, and .016 at T1, T2, and T3 respectively, for externalizing problems, and .031, .056 and .017 at T1, T2, and T3 respectively for parenting stress indicating that only a very small part of the variance could be explained by a class effect. Therefore, we decided to run the analyses without adjusting for clustering. Model fit was assessed with various fit indices, including robust chi-square with estimated degrees of freedom (df), comparative fit index (CFI); Bentler, 1990, root mean squared error of approximation (RMSEA; Byrne, 1998), and Tucker–Lewis index (TLI; Tucker & Lewis, 1973). Several models we tested were saturated (X² (0) = 0.00), therefore we did not report fit indices for these models.

Results

Descriptive Statistics
Correlations showed that internalizing and externalizing problems were moderately related to parenting stress across all time points (Table 1). Further, internalizing and externalizing problems were strongly correlated across time, indicating high stability of these problems. Parenting stress was also strongly correlated across time. Mental health was correlated positively and weakly to internalizing and externalizing problems, and moderately to parenting stress. Age correlated positively with internalizing problems at T1 and T2, indicating that for older children more internalizing problems were reported. However, age was correlated negatively with parenting stress, such that less parenting stress was reported for older children. At T1, a small correlation was found between externalizing problems and age; for older children, less externalizing problems were reported. Regarding gender, more externalizing problems were reported for boys.

Basic Growth Curves
Growth parameter estimates for all of the basic growth curve models are reported in Table 2. Fit statistics for the model investigating the shape and growth of internalizing problems were satisfactory (χ²(3) = 10.76, p = 0.013; CFI = 0.987; RMSEA = 0.041 (CI = 0.017–0.069); TLI = 0.957). No gender differences were found regarding initial level (intercept), change over time (slope), the variance around the initial level and change over time, or covariance among initial level and change over time, indicating that boys and girls do not differ in their development of internalizing problems (Δχ²(5) = 4.50, p > .05). Internalizing problems remain stable over time, indicated by a non-significant slope. Significant inter-individual differences in initial level, but not in
change of these problems were found, suggesting that children differ in their level of internalizing problems at baseline. The initial level of, and change in, internalizing problems were not related, suggesting that the level of internalizing problems is not associated with change in these problems.

Regarding externalizing problems, again an adequate fit to the data was found ($\chi^2(3) = 3.92, p = 0.27$; $CFI = 0.998$; RMSEA = 0.014 ($CI = 0.00$–$0.048$; $TLI = 0.994$). Several gender differences were found regarding initial level and change in externalizing problems over time. Boys showed higher levels of externalizing problems at baseline than girls ($\Delta\chi^2(1) = 5.43, p < .025$), and girls decreased faster in externalizing problems than boys did ($\Delta\chi^2(1) = 8.09, p < .005$). No gender differences were found regarding the variance around the initial level and change over time, or covariance among initial level and change over time ($\Delta\chi^2(3) = 4.86, p > .90$). Significant inter-individual differences in initial level, and in change of these problems were found, suggesting that children differ in both their level of externalizing problems at baseline and the decrease in these problems. The initial level of, and change in, externalizing problems were negatively related, such that children with higher levels of externalizing problems at baseline showed smaller decreases of externalizing problems over time.

Regarding parenting stress, a satisfactory fit to the data was found ($\chi^2(3) = 18.68, p = 0.000$; $CFI = 0.978$; RMSEA = 0.059 ($CI = 0.036$–$0.086$; $TLI = 0.928$). No gender differences were found regarding interrelations between internalizing problems and parenting stress ($\Delta\chi^2(3) = 1.54, p > .10$). Higher levels of baseline levels of internalizing problems were associated with higher levels of parenting stress and the course (i.e. stability) of internalizing problems was related to larger decreases in parenting stress. The level of internalizing problems was not

### Parallel Growth Curves

Correlations between growth parameter estimates of parenting stress and internalizing problems, and externalizing problems respectively are reported in Table 3. A satisfactory fit was found for the model investigating the interrelation of parenting stress and internalizing problems ($\chi^2(11) = 50.53, p = 0.000$; $CFI = 0.974$; RMSEA = 0.049 ($CI = 0.038$–$0.063$; $TLI = 0.933$). No gender differences were found regarding interrelations between internalizing problems and parenting stress ($\Delta\chi^2(3) = 1.54, p > .10$). Higher levels of baseline levels of internalizing problems were associated with higher levels of parenting stress and the course (i.e. stability) of internalizing problems was related to larger decreases in parenting stress. The level of internalizing problems was not
related to the decrease of parenting stress. Also, the level of parenting stress was not related to the course (i.e. stability) of internalizing problems.

Fit statistics for the model investigating the interrelation of parenting stress and externalizing problems were satisfactory ($\chi^2(11) = 35.34, p = 0.000; CFI = 0.984; RMSEA = 0.038 (CI = 0.025–0.053); TLI = 0.961). Higher baseline levels of externalizing problems were associated with higher baseline levels of parenting stress and decreases in externalizing problems were related to larger decreases in parenting stress. The level of externalizing problems was not related to the decrease of parenting stress, also, the level of parenting stress was not related to the decrease of externalizing problems. No gender differences were found regarding interrelations between externalizing problems and parenting stress ($\Delta\chi^2(3) = 6.210, p > .10$).

**Cross-lagged Models**

The model wherein the direction of effects of parenting stress on internalizing problems and vice versa was tested was saturated. Internalizing problems showed moderate to strong stability throughout childhood ($r = .27 - .51, p < .000$), as did parenting stress.

**Table 3** Correlations between the Growth Parameters in the Parallel Growth Curves

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internalizing Problems</strong></td>
<td><strong>Intercept</strong></td>
<td><strong>Slope</strong></td>
</tr>
<tr>
<td>Intercept</td>
<td>.049**</td>
<td>-.001</td>
</tr>
<tr>
<td>Slope</td>
<td>.003</td>
<td>.006**</td>
</tr>
<tr>
<td><strong>Externalizing Problems</strong></td>
<td><strong>Intercept</strong></td>
<td><strong>Slope</strong></td>
</tr>
<tr>
<td>Intercept</td>
<td>.022**</td>
<td>-.001</td>
</tr>
<tr>
<td>Slope</td>
<td>-.002</td>
<td>.002**</td>
</tr>
</tbody>
</table>

Note. **p < .01, * p < .05. Coefficients in parentheses represent the unstandardized correlations.

**Figure 1** Cross-lagged Model between Internalizing Problems and Parenting Stress.

Note. Numbers after variable names refer to data waves. **p < .01, * p < .05.

**Figure 2** Cross-lagged Model between Externalizing Problems and Parenting Stress, for boys. Note. Numbers after variable names refer to data waves. **p < .01, * p < .05.
found. Externalizing problems at T1 were related to parenting stress at T2 and externalizing problems at T2 were related to parenting stress at T3, such that more externalizing problems were related to more subsequent parenting stress. Also, parent-on-child effects were detected, such that parenting stress at T1 was related to externalizing problems at T2, and parenting stress at T2 was related to externalizing problems at T3. Thus, more parenting stress was related to more subsequent externalizing problems. For girls, externalizing problems at T1 were related to parenting stress at T3, such that more externalizing problems were related to more subsequent parenting stress. This relation was not found at the other time points. Parenting stress at T1 was related to externalizing problems at T2, such that more parenting stress was related to more subsequent externalizing problems. Again, this relation was not found at other time points.

Discussion

The present study examined the developmental patterns of parenting stress, internalizing and externalizing problems and the relations between parenting stress and internalizing and externalizing problems over time in children aged 4-9 from a large community sample. Additionally, we examined how parenting stress is related to each of these problem clusters. These associations were compared between boys and girls and we controlled for maternal mental health.

Development of problem behaviors

It is well known that externalizing problems tend to decrease during childhood (Maughan et al., 2008). These findings are replicated and extended by the current study, by differentiating the growth rates for boys and girls and by investigating inter-individual variability. In line with the literature, boys show higher levels of externalizing problems at age 4-7 and decrease slower than girls (Dishion & Patterson, 2006), while for internalizing problems no gender differences were found (Ford, Goodman, & Meltzer, 2003). Furthermore, this study showed that there are inter-individual differences in the initial level of externalizing and internalizing problems. Côté et al. focused on the preschool period and Sterba et al. focused on children aged 2-11. Possibly, differences in the course of internalizing problems are pushed by early developmental
changes in internalizing problems, and not so much by changes during middle childhood. This statement is speculative though and requires empirical testing. Further, more comprehensive instruments for assessing internalizing problems were utilized in these studies, Côté et al. used the Children’s Depressive and Anxiety Symptoms in preschool (DAS) and the Child Behavior Check List (CBCL) was used by Sterba et al., whereas the SDQ that we used was designed as a screening instrument, and hence, shorter. This may have led to more sensitive measurement of internalizing problems, and thus of more variation in reports of internalizing problems. If anything, our results call for the need for further validation of the course of internalizing problems. Regarding associations of the initial level of problems and its course, children with higher levels of externalizing problems at age 4-7 showed smaller decreases of externalizing problems over time. These results confirm the theory of antisocial development that more severe externalizing problems are associated with a worse outcome (Loeber, 1991). For internalizing problems, initial level of these problems was not related to its course.

Development of parenting stress and its relation to problem behaviors

Only one study investigated the developmental course of parenting stress (Williford et al., 2007). Our results concur with this study in the sense that parenting stress decreased over time and inter-individual differences in the initial level and in the course of parenting stress were found. These results mesh with extant literature showing that children’s independence increases in middle childhood, thereby decreasing the strain and demands on parents (Berk, 2012). Also, the initial level of parenting stress was negatively related to its course, implying that higher reports of parenting stress at baseline were related to smaller decreases in parenting stress over time. This may be explained by the role of cognitions about parenting stress (Lazarus, 1999), wherein these dysfunctional cognitions maintain perceptions of stress (Deater-Deckard, Smith, Ivy, & Petrill, 2005). Thus, when parenting stress is present at some point, it is likely that dysfunctional cognitions maintain perceptions of stress, thereby reducing the decrease in perceived.

The current study expanded on Williford et al. (2007), by investigating how the development of parenting stress was related to both internalizing and externalizing problems. As expected, parenting stress and internalizing and externalizing problems were interrelated at baseline (e.g., Crnic et al., 2005; Rodriguez, 2011). The decrease in parenting stress was not affected by the initial level of internalizing and externalizing problems, which was contrary to our hypothesis that higher levels of parenting stress are associated with smaller decreases in problem behavior (Deater-Deckard, 2004). These findings suggest that other factors influence the course of parenting stress. Possibly, socioeconomic stressors, such as economic hardship, affect aspects of personality, by increasing ineffective coping styles, (Crnic et al., 2005; Warfield, 2005), which in turn affect the course of parenting stress. However, the course of parenting stress was related to the course of internalizing and externalizing problems, suggesting that these constructs co-evolve during childhood. Together, these findings suggest a complex interplay between factors related to the child and to the parent, which is in accordance with a transactional model (Sameroff, 1975; Cicchetti, 2006; Minuchin, 1985).

As for directionality, consistent patterns in parenting stress and externalizing problems were found for boys above and beyond stability of these problems and while taking into account concurrent relations. For girls, the pattern was less consistent but still in accordance with a bidirectional model, as parenting stress at baseline impacted externalizing problems one year later, and externalizing problems at baseline impacted parenting stress two years later. These results fit well with the theoretical notion that both parents and children affect each other’s development, as described in the bidirectional model (Sameroff, 1975; Cicchetti, 2006; Minuchin, 1985). However, regarding internalizing problems, no support was found for the child effects model, as internalizing problems did not affect subsequent parenting stress above and beyond stability of these problems and while taking concurrent relations into account. Some support was found for a parent effects model, as parenting stress at baseline was associated with internalizing problems one year later. These results diverge from those found in the growth models, and may be explained by analytic principles. While growth models take inter-individual differences into account, in cross-lagged models only the group level is included. Moreover, time is used in a different way in these analyses, as in cross-lagged analyses each time point is used separately, whereas in growth models change is modeled as change across time.

Limitations and future directions

First, given the consistent empirical finding that informants tend to disagree regarding problem behavior and parenting (De Los Reyes & Kazdin, 2005; Taber, 2010), the results of this study may be informant specific, that is specific to mothers. As such, our results should not be generalized to fathers and to other informants such as children. Also, the possibility that our results reflect shared method variance can not be ruled out (De Los Reyes & Kazdin, 2005). Future studies should include reports of multiple informants. Interestingly, such a framework may actually help to explain how maladjustment develops (De Los Reyes, 2012), instead of making the results harder to interpret. For example, parent-child discrepancies regarding parenting stress may represent features of the parent-child relationship, which in turn affect child outcomes (Goodman, De Los Reyes, & Bradshaw, 2010). Moreover, analytical techniques enable partitioning out non-shared variance among informant’s reports, and testing whether these reflect unique information, instead of treating this variance as error (e.g., Bartels, Boomsma, Hudziak, van Beijsterveldt, & van Oordt, 2007). Relatedly,
given the large sample size, only questionnaires were used in this study. Combining questionnaire data with observations would provide a more thorough and valid test of interrelations between parenting stress and problematic behaviors. Second, as stated above, our sample is normative and relatively low risk. This means that our results should not be generalized to high risk or clinical samples per se. However, the spectrum hypothesis states that differences between normative and clinical samples primarily lie in mean level differences and in strength of associations, but not in structural or qualitative differences (Costa & Widiger, 2002; Van Leeuwen, Mervielde, De Clerq, Fruyt, 2007). By including both normative and clinical samples in future studies, this hypothesis can be tested regarding relationships between parenting stress and problem behaviors.

Conclusions

This study employed a developmental perspective to investigate the interrelatedness between parenting stress and internalizing and externalizing problems. Given the limited studies investigating these interrelations dynamically, the current study is an important contribution to the extant literature. We showed that parenting stress is not a static construct and decreases during early childhood, similarly for boys and girls. In addition, we showed that internalizing problems remain stable in early childhood, and replicated findings regarding decreasing externalizing problems. Finally, parenting stress and internalizing and externalizing co-evolve, thereby mutually influencing each other. Regarding internalizing problems, however, child effects were not found. This finding calls for the need for future studies to further disentangle the interrelatedness of parenting and internalizing problems.

References


Part II

Child Perceptions of Problem Behaviours and Parenting Behaviours
The Berkeley Puppet Interview: A Screening Instrument for Measuring Psychopathology in Young Children

Published as:
Abstract

While child self-reports of psychopathology are increasingly accepted, little standardized instruments are utilized for these practices. The Berkeley Puppet Interview (BPI) is an age-appropriate instrument for self-reports of problem behavior by young children. Psychometric properties of the Dutch version of the BPI will be reported, specifically, test-retest reliability, intra-class correlations, congruent and concurrent validity. In a sample of 300 children (M_{age}=7.04 years, SD=1.15), the BPI was administered twice, with a one-year interval. Parents and teachers filled out questionnaires about their children’s problem behavior. Findings from the analyses indicate that the BPI subscales have sufficient test-retest reliability and can be reliably coded. Furthermore, findings suggest adequate congruent validity. More support for concurrent validity is found among externalizing problems in comparison to internalizing problems. With regard to the present study, the BPI seems to have adequate psychometric properties. As such, the BPI enables interviewing young children about their psychopathology-related symptoms in a standardized way. The BPI could be applied in clinical practice as a complement to the diagnostic cycle, allowing children’s self-reports to play an increasingly important role.

Introduction

Problem behavior often develops at a young age. A considerable number of children suffer from mental health problems. Prevalence figures show that between three and eighteen percent of children exhibit symptoms of psychopathology (Carter, Wagmiller, Gray, McCarthy, Horwitz, & Briggs-Gowan, 2010; Costello, Mustillo, Erkanli, Keeler, & Angold, 2003). Externalizing problems, such as oppositional defiant behavior, antisocial behavior, and attention difficulties, as well as internalizing problems, including separation anxiety, anxiety, and depressive symptoms, are most common in young children (Egger & Angold, 2006; Klein, Dougherty, & Olino, 2005; Lavigne, LeBailly, Hopkins, Gouze, & Binns, 2009). In addition, co-morbidity is quite common, especially with regard to young children (Lavigne et al., 2009; Scheeringa & Zeanah, 2008).

It is important to be able to examine psychopathology at a young age, since high degrees of aggressive and oppositional behavior may become permanent and develop into chronic patterns of externalizing and psychopathological behavior at a later age (Reef, Diamantopoulous, van Meurs, Verhulst, & van der Ende, 2010). Problem behavior is associated with increased risks of poor academic, social and occupational performance, deteriorated physical and mental health, and substance use (Ansary & Luthar, 2009; Bayer, Rapee, Hiscock, Ukoumunne, Mihalopoulos, & Wake, 2011; Fergusson, Horwood, & Ridder, 2005; Kim, Guerra, & Williams, 2008; Morcillo et al., 2011; O’Neill, Conner, & Kendall, 2011). When assessed early in development, interventions may contribute to the reduction of aggressive, oppositional and other externalizing behaviors, before these negative behavioral patterns become integrated into the child’s personality (Hill, Lochman, Coie, & Greenberg, 2004).

Several factors have contributed to the phenomenon that, in both research and clinical practice, the emphasis is on externalizing rather than internalizing problems. Probably, one major reason behind this is that externalizing behavior is easier to observe than internalizing behavior. Externalizing behaviors, such as tantrums and resistance against rules, are outwardly directed, generally troublesome for the environment, and often provocative in terms of negative feelings (Rubin & Mills, 1990). On the other hand, internalizing problems are intra-individual in nature, inwardly directed, and more easily shielded from the environment by the child (Luby, Si, Belden, Tandon, & Spitznagel, 2009). These behaviors attract less attention and cause fewer problems for the child’s environment. Of course, a child may still experience such internalizing problems and suffer from them. Indeed, research shows that even young children report on internalizing problems (Luby, 2010), and that these problems are related to negative developmental outcomes later in life, including recurrent depressive episodes, poor school performance, impaired functioning of peer and family relationships, and an increased risk of suicide (Bhatia...
The fact that internalizing problems at a young age are predictive of problems at a later age, stresses the need of early intervention (Bayer & Sanson, 2003).

Yet, while 50% of children expressing externalizing behaviors receive help, this is true for only 20% of children suffering from internalizing problems (Menikagas et al., 2011). Some researchers suppose that internalizing problems are generally better recognized by children themselves than by other informants (Achenbach, McConaughy, & Howell, 1987). In one respect, it is possible that an informant’s background distorts his/her perception of a child’s behavior, particularly when the behavior is more ambiguous, as is the case with internalizing problems (Kroes, Veerman, & De Bruyn, 2003). For example, personality characteristics such as hostility and inadequate interpersonal sensitivity, are associated with reporting on internalizing problems. In another respect, it is likely that children behave differently in several environments (e.g., at home versus at school), which ensures that information derived from different informants is related to the specific context by definition. Hence, the problem with obtaining information from different informants is that these perceptions are context specific and biased by personal backgrounds (De Los Reyes & Kazdin, 2005). Alongside conventional screening instruments that are used during the problem analysis phase in clinical practice, including the CBCL/TRF and SDQ (Achenbach & Ruffle, 2000; Goodman, Ford, Simmons, Catward, & Meltzer, 2000), it seems worthwhile to pay attention to the possibility of adopting instruments that refer to the child as an informant. This is in accordance with the so-called ‘multi-informant approach’, in which it is recommended to take into account context (i.e., at home and elsewhere), and perspective (i.e., self and other), when selecting informants (Kraemer et al., 2003). By using self-report instruments, the risk of under-reporting of internalizing problems may be reduced and a more comprehensive picture of the existing problems will arise (Kraemer et al., 2003).

Screening instruments use self-reports of young children to a minor extent. Young children are not always considered reliable informants of their own behavior (Mutsaers, 2009; Scheerings & Haslett, 2010). Children’s vocabulary and cognitive development may affect their understanding of questions and interfere with the duration of administration (Arseneault, Kim-Cohen, Taylor, Caspi, & Moffit, 2005). Furthermore, it is doubted whether children are capable of self-perception, as this concept is related to cognitive development. Moreover, young children are very sensitive to suggestion, which makes interviewing children a challenge and requires specific interviewing skills. Still, already in the 80’s, Harter (1982) showed that children from the age of eight can meaningfully differentiate between various competence scales (cognitive, social, and physical competence, and general self-esteem). Measelle, Ablow, Cowan, and Cowan (1998) stated that children’s self-perceptions can indeed be reliably measured by using an age-appropriate instrument. In clinical practice it is also known that children from six years can be interviewed as a part of the diagnostic cycle (Van Leeuwen, 2002), thereby adding unique information to the diagnostic process. In the last few years, children’s self-reports are valued increasingly (Arseneault et al. 2005; Ialongo, Edelsohn, & Kellam, 2001; Luby, Belden, Sullivan, & Spitznagel, 2007). Specific self-report questionnaires are available for children from eight years onwards, such as the Child Depression Inventory (CDI; Kovacs, 2001), Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher, Brent, Chiappetta, Bridge, Monga, & Baughner, 1999), and Perceived Competence Scale for Children (PCSC; Harter, 1982). However, in practice, there is no screening instrument available in the Netherlands, that uses children younger than eight years old as informants for the assessment of their psychopathology. The Berkeley Puppet Interview (BPI; Measelle et al., 1998; Morris et al., 2002) is an interactive interviewing technique, developed in the USA and designed to elicit perceptions of 4.5 to 8-year-olds in an age-appropriate way. During the BPI, children are interviewed by two hand puppets in order to simulate a conversation between three peers. Each time, these two hand puppets make opposing statements. For example, one puppet indicates: “I am a sad child”, whereas the other puppet states: “I am not a sad child”. Then, they ask the child together: “How about you?” Influencing the child in the direction of the question that is asked by the interviewer is thus largely avoided.

In previous studies the BPI has proven to be reliable and valid (Ablow et al., 1999; Arseneault et al., 2005; Luby et al., 2007; Measelle et al., 1998; Morris et al., 2002; Ringoot et al., 2013). However, only one of these studies used longitudinal data and the sample of this study was rather small with less than 100 participants (Measelle et al., 1998). In addition, recent former studies investigated specific problem clusters of the BPI, such as conduct problems or depression (Arseneault et al. 2005, Luby et al., 2007), with one exception (Ringoot et al., 2013), our aim is to investigate the BPI as a whole. Further, more research into the BPI’s psychometric properties may facilitate its use in clinical practice. As such, the BPI may be suitable for embedding into the diagnostic cycle. Clinicians naturally conduct interviews with children, and the BPI allows doing so in a standardized manner, without disregarding particular case-dependent questions. In addition, it is an age-appropriate instrument of which the administration will take less time than a diagnostic interview. Recently, the BPI was used as a research instrument as part of two large-scale studies in the Netherlands: the Kid in Zicht study (Stone, Giletta, Brendgen, Otten, Engels, & Janssens, 2013a), and the Generation R study (Jaddoe et al., 2012). Kind in Zicht is a longitudinal research project on incipient emotional and behavioral problems in young children (Stone et al., 2013a). Generation R involves research into early influences on growth and development within a longitudinal multi-ethnic birth cohort (Jaddoe et al., 2012). For the BPI to be used in these studies, a Dutch version was developed in collaboration with the developers of the instrument.
In the present article, we introduce the Dutch version of the BPI as a useful instrument complementary to the diagnostics in the field of psychopathology, and we examine the test-retest reliability, and the congruent and concurrent validity of the BPI in the Kind in Zicht study. We expected the Dutch version of the BPI – like the American version – to be a reliable and valid instrument for self-reports of psychopathology in young children.

**Method**

**Sample and Procedure**

In this study, 300 children were interviewed during the first measurement (T1). One child was excluded due to missing data and another child because she was over eight years old. One year later (T2), 288 of these children (96%) were re-interviewed, of whom one was excluded because of her advanced age. This resulted in a sample of 298 children at T1, and 287 children at T2. Of these participating children, 50% was male and the mean age was 6.95 years (SD=1.13; range 5-8 years). The majority of the children was of Dutch origin (97.4%) and grew up in a two-parent family (92.2%).

Teachers (T1 n=282, T2 n=245) and parents (T1 n=289, T2 n=269) completed questionnaires about the children at both time points. In addition, teachers (n=287) and parents (n=287) completed a questionnaire about the children one year before the interviews took place, and this measurement point is referred to as T0. At T0, the teachers’ mean age was 36.57 years (SD=10.43), and 93.9% of them was female. The parents who filled out the questionnaires were on average 38.29 years old (SD=3.88), and 92.9% of them was female. Over half of the parents were highly educated (54.8%), 37.3% had an intermediate education level, and 6.6% lower education. Slightly over 1% received some other type of education.

For the present study, longitudinal data (2011(T1)-2012(T2)) from the Kind in Zicht project were used (Stone et al., 2013a), which was approved by the committee on ethics. Within this project, information was collected about the individual children, using multiple informants. Informed consent from the children’s parents was obtained. Each year, the BPI was administered to the children by five certified master students or researchers. They all completed a training course in which the interviewing techniques of the BPI were extensively practiced. Subsequently, they each conducted eight practice interviews, and were then evaluated. The interviews were administered at primary schools in January and February of 2011 and 2012. Children were interviewed in an empty classroom to ensure confidentiality. Interviews were videotaped and after completion, the children received a pair of stickers to thank them for their participation.

**Measures**

**BPI.** The Berkeley Puppet Interview (BPI; Measelle et al., 1998) is an interactive and age appropriate interviewing technique, designed to elicit self-perceptions in 4.5 to 8 year-olds. During the interview, children were asked questions by two identical hand puppets: Iggy and Ziggy. Prior to the interview, the puppets introduced themselves and explained in a playful way how the interview is carried out. Using three practice items, the interviewer assessed whether the procedure was clear to the child, and continued with the actual interview or repeated the practice items until the procedure was clear. An example of such a practice item is: Puppet 1: ‘I like chocolate’, Puppet 2: ‘I do not like chocolate. How about you?’. Throughout the interview, the puppets exchanged opposing statements and then asked the child: ‘How about you?’. The puppet with which the child agreed repeated the response, thereby confirming the child’s answer.

![Figure 1](image1.png)

**Figure 1** Pictures of the Berkeley Puppet Interview.

After administration of the interviews, the children’s answers were coded by trained observers on a 7-point scale (see Figure 2). Answers that reflected the absence of psychopathology were coded as either 5, 6, or 7, depending on possible
amplifications or attenuations in the child’s response. Code 7 comprised the strongest absence of psychopathology (e.g., ‘I am never a sad child’), whereas code 6 meant a neutral absence (e.g., ‘I am not a sad child’), and code 5 represented a hesitant response (e.g., ‘Usually, I am not a sad child’). On the other side of the spectrum, code 1, 2, or 3 reflected the presence of psychopathology. Code 1 stood for a strong presence (e.g., ‘I am always a sad child’), while code 2 represented a neutral response (e.g., ‘I am a sad child’), and code 3 was equivalent to a hesitant response (e.g., ‘Usually, I am a sad child’). When a child was unable to choose between the two statements, this response was coded as 4. In order to test the reliability of the coding, 15% of the interviews were double-coded.

The BPI includes 8 subscales (i.e., the symptom scales), that constitute the basis for two overall scales: internalizing problems and externalizing problems. The internalizing problems scale comprises three subscales: depression (7 items; e.g., ‘I am a sad child/I am not a sad child’), anxiety (7 items; e.g., ‘I do have many bad dreams/I do not have many bad dreams’), and separation anxiety (6 items; e.g., ‘When I am at school, I miss my mom or dad/When I am at school, I do not miss my mom or dad’). We used the internalizing problems scale, as well as the separate symptom scales. The externalizing problems scale also comprised three subscales: oppositional defiant behavior (6 items; e.g., ‘Sometimes I curse, or I use bad language/I do not curse, or I use bad language’), behavioral problems (9 items; e.g., ‘Sometimes I act cruel towards animals/I do not act cruel towards animals’), and aggression and hostility towards peers [from here referred to as aggression] (6 items; e.g., ‘I often fight with other children/I do not fight with other children’). In addition, two subscales focus on relationships with peers: acceptance and rejection by peers [from here referred to as acceptance/rejection] (5 items; e.g., ‘Other children ask me to play along/Other children do not ask me to play along’), and being bullied (4 items; e.g., ‘Children hit me, or beat me up/Children do not hit me, or beat me up’). The negative and positive statements were presented in a random order. No Cronbach’s alpha’s will be reported regarding the BPI, since the interview is considered an index scale instead of a Likert scale, making it unsuitable for calculating this reliability coefficient (Stone, Otten, Janssens, Soenens, Kuntsche, & Engels, 2013b). The interrater reliability is reported in the results section.

SDQ. The Dutch parent and teacher version of the Strengths and Difficulties Questionnaire (SDQ) was used to assess internalizing and externalizing problems (Widenfelt, Goedhart, Treffers, & Goodman, 2003). The subscales measuring emotional problems (e.g., often unhappy, down-hearted or tearful) and behavioral problems (e.g., often lying or cheating) each consist of five items. Parents or teachers judged children on a 3-point scale, from 0 (not true) to 2 (very true). The scoring manual is available online (www.sdqinfo.com). In the Kind in Zicht study, the psychometric properties of the SDQ were adequate, as described elsewhere (Stone et al., 2013b).

CBCL/TRF. The Dutch versions of the Child Behavior Check List (CBCL) and Teacher Report Form (TRF) were also used (at T0) to measure internalizing and externalizing behavior, as reported by parents and teachers (Achenbach & Rescorla, 2000; Achenbach & Rescorla, 2001; Verhulst, Van der Ende, & Koot, 1997). The C-TRF and C-CBCL are intended for children aged 1.5 to 5 years and contain 100 items; the TRF and CBCL are intended for 5 to 18 year-olds and contain 118 items. The C-TRF and TRF were filled out by teachers, whereas the C-CBCL and CBCL were filled out by parents. Items were scored on a 3-point Likert scale, from 0 represents ‘not true’, and 2 stands for ‘very true or often true’. Three scales (i.e., somatic symptoms, anxious-depressed, and withdrawn) were combined in order to constitute the internalizing scale. Combining two scales (i.e., violation of rules and aggressive behavior) resulted in the externalizing scale. The psychometric properties of this instrument in the Kind in Zicht study were again adequate (Stone et al., 2013b).

Strategy for Analysis
First, descriptive statistics that provide insight into the level of psychopathology for the whole sample will be shown, disaggregated for gender and age group (4-5 and 6-7 years). Besides, an independent t-test was conducted to test whether the mean scores of boys and girls, and younger and older children, respectively, differ statistically. Originally, the BPI is scored in such a way that higher scores reflect lower levels of psychopathology. In our opinion, this is somewhat confusing. For the sake
of clarity regarding the interpretation, the scores were therefore coded the other way around (i.e., 1 becomes 7, and vice versa), such that higher means reflected higher levels of problem behavior. These reversed scores were used for calculating means and standard deviations.

Subsequently, the reliability of the BPI codes was examined using intra-class correlations and test-retest correlations. The intra-class correlation coefficient (ICC) was calculated to determine the reliability between two coders per BPI subscale. The higher the ICC, the more reliable the coding, where a score of 1 represents absolute agreement. ICC values of > .80 are considered good and values > .75 are considered excellent (Cicchetti, Koenig, Klin, Paul, & Sparrow, 2011). Pearson correlations were used for calculating test-retest correlations. These test-retest correlations were calculated for the entire group, and for gender and age separately.

In terms of validity, congruent validity was examined first by mutually correlating the BPI subscales. Additionally, concurrent validity was defined by correlating the BPI outcomes with the outcomes of the other questionnaires; again using Pearson correlations. When comparing the BPI with the SDQ and CBCL, the BPI subscales were ranged under two headings; the internalizing problems scale and the externalizing problems scale. These were compared with the emotional and behavioral problems scale of the parent and teacher versions of the SDQ. The CBCL also used an internalizing and externalizing problems scale, that was completed by both parents (CBCL) and teachers (TRF). Because of the ages of a restricted group of children, alternative versions were deployed; the C-CBCL and the C-TRF. In order to clearly show the possible similarities and differences between the BPI and CBCL, the standardized T-scores of the CBCL and C-CBCL, and those of the TRF and the C-TRF were combined.

**Results**

**Descriptive Statistics**

The descriptive statistics of the BPI subscales appear in Table 1. The mean scores on the subscales were low. T-tests for paired observations showed that the mean scores of depression, separation anxiety, anxiety, behavioral problems, and being bullied, declined from T1 to T2. In addition, it was tested whether mean differences regarding age and gender at T1 and T2 were present. The t-test for gender at T1 showed that there were statistically significant mean differences for separation anxiety (t(286) = -2.25, p < 0.05), aggression (t(289) = 3.56, p < 0.01), and acceptance/rejection (t(284) = 2.04, p < 0.05), but not for depression, anxiety, behavioral problems, oppositional defiant behavior, and being bullied. The mean scores of boys on the aggression and acceptance/rejection subscales were higher than those of girls, while girls scored higher on separation anxiety than boys. At T2, the t-test for gender was statistically significant for the subscales separation anxiety (t(279) = -3.37, p < 0.05), oppositional defiant behavior (t(280) = 3.02, p < 0.05), behavioral problems (t(280) = 2.07, p < 0.05), aggression (t(279) = 3.96, p < 0.01), acceptance/rejection (t(280) = 2.49, p < 0.05), and being bullied (t(279) = 2.39, p < 0.05), but not for depression and anxiety. Mean scores of boys at T2 were higher than those of girls on the subscales oppositional defiant behavior, behavioral problems, aggression, acceptance/rejection, and being bullied, whereas girls reported higher scores on separation anxiety than boys. In conclusion, boys generally reported more externalizing problems than girls at both time points.

As regards the t-test for age, mean scores for depression (t(282) = 2.46, p < 0.05) and acceptance/rejection (t(276) = 2.22, p < 0.05) were found to be higher for younger children as opposed to older children at T1. At T2, younger children also reported more symptoms of depression (t(273) = 2.76, p < 0.01), as well as aggression (t(272) = 2.12, p < 0.05), and they indicated to be bullied more than older children (t(272) = 3.77, p < 0.01).

**Intra-class Correlations**

The following ICC’s were obtained for the separate subscales, for T1 and T2 respectively: depression (74, 86), anxiety (70, 80), separation anxiety (70, 83), oppositional defiant behavior (66, 71), behavioral problems (81, 66), aggression (78, 77), acceptance/rejection (82, 82), and being bullied (74, 88). These correlations indicated that the BPI subscales can be reliably coded by multiple coders.

### Table 1 Descriptive statistics of the BPI subscales at T1 and T2

<table>
<thead>
<tr>
<th>Subscale</th>
<th>T1 N = 291-297</th>
<th>T2 N = 286-287</th>
</tr>
</thead>
<tbody>
<tr>
<td>M (SD)</td>
<td>Range</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Depression</td>
<td>2.61 (.62)</td>
<td>1.71-5.00</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3.10 (.83)</td>
<td>1.71-5.67</td>
</tr>
<tr>
<td>Separation anxiety</td>
<td>3.33 (.91)</td>
<td>1.83-6.67</td>
</tr>
<tr>
<td>Oppositional/defiant</td>
<td>2.67 (.61)</td>
<td>1.33-4.80</td>
</tr>
<tr>
<td>Behavioral problems</td>
<td>2.56 (.64)</td>
<td>1.22-5.11</td>
</tr>
<tr>
<td>Aggression</td>
<td>2.32 (.57)</td>
<td>1.33-5.33</td>
</tr>
<tr>
<td>Acceptance/rejection</td>
<td>2.67 (.84)</td>
<td>1.60-6.00</td>
</tr>
<tr>
<td>Being bullied</td>
<td>3.01 (1.06)</td>
<td>1.50-6.25</td>
</tr>
</tbody>
</table>

Note. * p < 0.05, ** p < 0.01
Test-retest Reliability
In Table 2, the results with regard to test-retest reliability, with a time interval of one year, are presented. These showed that, overall, the psychopathology self-reports as provided by the children were rather stable. Boys appeared to report somewhat less stable than girls, in terms of oppositional defiant behavior, behavioral problems, and being bullied. Moreover, the correlations regarding depression, separation anxiety, acceptance/rejection, and being bullied were less pronounced in young children than in older children. The test-retest reliability of these scales thus increased with age.

Congruent Validity
As is apparent from Table 3, the BPI subscales correlated significantly at T1 and T2. The correlations were weak to moderate, and the pattern of correlations was as expected: the reports of certain types of problem behaviors were associated with the reports of other types of problem behaviors (e.g., anxiety was correlated with depression). The internalizing subscales depression, separation anxiety, and anxiety, correlated weakly with the externalizing subscales oppositional defiant behavior, behavioral problems, and aggression. The correlations between the internalizing subscales themselves were stronger, especially between anxiety and depression, and anxiety and separation anxiety. Furthermore, oppositional defiant behavior, behavioral problems, and aggression correlated relatively strongly with one another. Acceptance/rejection correlated predominantly with depression and oppositional defiant behavior, and to a lesser extent with behavioral problems, aggression, and anxiety. The subscale being bullied was correlated with all other subscales. In summary, various problem behaviors were meaningfully intercorrelated within this young age group.

Concurrent Validity
The externalizing subscales of the BPI and the SDQ were correlated at T1 and T2, concerning both parents and teachers (see Table 4). The more externalizing problems the children reported, the more behavioral problems parents and teachers reported likewise. It is noteworthy that the internalizing subscales of the BPI and the SDQ correlated to a lesser extent than the externalizing subscales. In order to explain this difference, the individual internalizing BPI subscales (i.e., anxiety, depression, and separation anxiety), were correlated to the SDQ emotional problems scale score. Depression, separation anxiety, and anxiety were uncorrelated with emotional problems as reported by teachers at T2, but depression did: \( r(238) = .21, p < .01 \).

As for the parents as informants, it was noticed that whereas at T1 the internalizing BPI scale was correlated with emotional problems, it was no longer at T2. Next, idem, the separate BPI subscales were correlated to the SDQ emotional problems scale score. At both time points, no correlation was found between separation anxiety and emotional problems (T1: \( r(286) = .08, n.s. \); T2: \( r(255) = .02, n.s. \)) and between anxiety and emotional problems (T1: \( r(276) = .08, n.s. \); T2: \( r(255) = .02, n.s. \)). Depression was found to be associated with emotional problems at both T1 and T2 (T1: \( r(286) = .12, p < .05 \); T2: \( r(256) = .17, p < .01 \)). From these results, we can conclude that children’s self-reports of depression corresponded to some extent to the emotional problems reports by teachers and parents; the more emotional problems teachers and parents reported, the more depression children reported. However, children’s self-reports of anxiety and separation anxiety did not correspond to teachers’ and parents’ reports of emotional problems.

The BPI subscales measured at T1 have also been compared with the CBCL/ TRF scale scores at T0. Children’s self-reported internalizing problems did not correlate with parent’s and teachers’ reported problems (\( r(278) = -.00, n.s. \); \( r(286) = -.02, n.s. \), respectively). However, the correlations between children’s self-reports and the reports of their parents (\( r(279) = .20, p < .01 \) and teachers (\( r(287) = .14, p < .05 \)) on externalizing problems were significant. Children’s reports regarding internalizing problems were not correlated with the reports of parents and teachers about the children’s behaviors in the previous year, while children’s reports regarding externalizing problems were.
Discussion

At present, no standardized instrument is available in the Netherlands for measuring self-perceptions of problem behavior in young children (Mutsaers, 2009). This is problematic, since it is known that there may be great differences in reports of parents and teachers about children’s behaviors (De Los Reyes & Kazdin, 2005). As a consequence, certain problem behaviors may not be recognized. Therefore, it is important that attention is paid to self-reports of problem behavior by young children.

In this article, the Dutch version of the Berkeley Puppet Interview (BPI) was presented, which is a standardized and age-appropriate instrument for interviewing young children about their self-perceptions of problem behaviors. In addition, several psychometric properties of the BPI were presented.

We expected that the results regarding reliability and validity would be consistent with earlier research into the BPI. The results suggest that the BPI scales can be sufficiently reliably coded, that the subscales are correlated after one year, and that the subscales are meaningfully intercorrelated, which indicates congruent validity.

The analyses concerning the intra-class correlation coefficients and test-retest reliability imply that the BPI is a consistent, reliable interviewing method. Though, it should be noted that the intra-class correlation for oppositional defiant behavior were somewhat lower. The interpretation of the results of this subscale should be interpreted with some caution. Still, even after a one-year interval, during which, of course, not only reliability was assessed, but also development, there appeared to be clear

Table 3 Correlations among the BPI subscales at T1 and T2

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Depression</td>
<td>- .34**</td>
<td>- .38**</td>
<td>- .19**</td>
<td>- .32**</td>
<td>- .20**</td>
<td>- .13**</td>
<td>- .15**</td>
<td>- .11**</td>
</tr>
<tr>
<td>2 Separation anxiety</td>
<td>- .40**</td>
<td>- .31**</td>
<td>- .31**</td>
<td>- .43**</td>
<td>- .38**</td>
<td>- .26**</td>
<td>- .19**</td>
<td>- .17**</td>
</tr>
<tr>
<td>3 Anxiety</td>
<td>- .18**</td>
<td>- .13**</td>
<td>- .05</td>
<td>- .03</td>
<td>.03</td>
<td>.08</td>
<td>.14**</td>
<td>.28**</td>
</tr>
<tr>
<td>4 Oppositional/defiant</td>
<td>- .28**</td>
<td>- .17**</td>
<td>- .05</td>
<td>- .03</td>
<td>.03</td>
<td>.08</td>
<td>.14**</td>
<td>.28**</td>
</tr>
<tr>
<td>5 Behavioral problems</td>
<td>- .18**</td>
<td>- .19**</td>
<td>- .30**</td>
<td>- .37**</td>
<td>- .35**</td>
<td>- .27**</td>
<td>- .19**</td>
<td>- .15**</td>
</tr>
<tr>
<td>6 Aggression</td>
<td>- .18**</td>
<td>- .19**</td>
<td>- .11</td>
<td>.08</td>
<td>.09</td>
<td>.18**</td>
<td>- .03</td>
<td>- .03</td>
</tr>
<tr>
<td>7 Acceptance/rejection</td>
<td>- .18**</td>
<td>- .19**</td>
<td>- .11</td>
<td>.08</td>
<td>.09</td>
<td>.18**</td>
<td>- .03</td>
<td>- .03</td>
</tr>
<tr>
<td>8 Being bullied</td>
<td>- .18**</td>
<td>- .19**</td>
<td>- .11</td>
<td>.08</td>
<td>.09</td>
<td>.18**</td>
<td>- .03</td>
<td>- .03</td>
</tr>
</tbody>
</table>

Note. Numbers above the diagonal T1 (N = 291-297), numbers below the diagonal T2 (N = 286-287)

Table 4 Correlations among the BPI subscales and the SDQ T1 and T2 scale scores

<table>
<thead>
<tr>
<th></th>
<th>Parent Emotional Problems</th>
<th>Parent Behavioral Problems</th>
<th>Teacher Emotional Problems</th>
<th>Teacher Behavioral Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPI Internalizing</td>
<td>.14*</td>
<td>.05</td>
<td>.09</td>
<td>.10</td>
</tr>
<tr>
<td>BPI Externalizing</td>
<td>.11</td>
<td>.29**</td>
<td>-.03</td>
<td>.25**</td>
</tr>
<tr>
<td>SDQ T1 Emotional Problems</td>
<td>.08</td>
<td>.06</td>
<td>.17**</td>
<td>.08</td>
</tr>
<tr>
<td>SDQ T1 Behavioral Problems</td>
<td>.23**</td>
<td>.23**</td>
<td>.27**</td>
<td>.27**</td>
</tr>
</tbody>
</table>

Note. * p < 0.05, ** p < 0.01
patterns in the behaviors children report. The test-retest coefficients are not as high as typically found in studies that focus on adults, but are similar to other studies investigating the BPI’s psychometric properties (Measelle et al., 1998). Furthermore, theoretically speaking, it was to be expected that the BPI subscales were meaningfully interrelated. This indicates that the BPI seems to measure the constructs that are intended to be measured. However, for determining congruent validity, it is also necessary that the BPI will be compared to external measures, such as standardized tests that assess school performance. Although children are sometimes still not considered reliable informants of their own problems (Mutsaers, 2009; Scheeringa & Haslett, 2010), the results of this study seem to indicate the opposite. This is in line with other studies that have been conducted into the BPI (Arsenault et al., 2005; Luby et al., 2007; Measelle et al., 1998), and with recommendations to clinicians, that children from the age of 6 years can be interviewed as part of the diagnostic cycle (Van Leeuwen, 2003). The comparison of the BPI with the SDQ and CBCL/TRF, shows that differences between reports of multiple informants are indeed great. It is important to note that comparing scores on the BPI on the one hand, and the SDQ and CBCL/TRF on the other hand is difficult, given the nature of the instruments; an interviewing technique versus a questionnaire. In spite of this difference in method, the correlations between comparable concepts measured using the BPI and SDQ or CBCL/TRF remain weak.

This phenomenon, ‘informant disagreement’, is a well-known issue when comparing reports from multiple informants (De Los Reyes & Kazdin, 2005). As expected, the agreement was greater in terms of externalizing behavior, than with respect to internalizing behavior, although the agreement on externalizing behavior was also very low. These results underscore that reports of problem behavior by parents and teachers cannot simply be regarded as corresponding to children’s perceptions (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005), particularly when it comes to reporting internalizing problems, where agreement between children and parents and teachers was very limited (Achenbach et al., 1987). These results also imply that child reports provide important information additional to the process of information gathering in the problem analysis phase. In this respect, the BPI could be a useful instrument. Based on the current state of research into the BPI, however, clinicians are recommended to also keep in mind the limitations of the BPI, when using this instrument. It is not recommended to use the BPI as a single instrument, but it seems suitable for gaining more insight into certain symptoms and for confirming or rejecting hypotheses regarding a child’s symptoms. In addition to the BPI, another promising instrument is available for children aged 6 to 11 years old: the Dominic Interactive (DI; Valla, 2000; Kuijpers, Otten, Krol, Vermulst, & Engels, 2013). The DI is a structured digital questionnaire that assesses the most common internalizing and externalizing problems in children. It takes into account the child’s developmental level, by means of supporting the questions by visual and auditory stimuli. The item is both displayed through an image of the problem situation, and made audible by being read out loud by the program.

Limitations and future directions

The present study showed that the BPI has adequate psychometric properties, although we believe that more research into the internal structure of the BPI is necessary and highly recommended for further research. A recent study did confirm the internal structure of the BPI and reported Cronbach’s alpha’s for the subscales (Ringoot et al., 2013). Yet, a thorough test of the internal structure of the BPI is hampered by the bimodal frequency distribution, and in our opinion as such, not suitable for the execution of conventional reliability analyses, such as calculating Cronbach’s alpha and testing the factor structure. The BPI thus appears to be a sound and useful instrument which could be used in child and youth care. Still, it is important that, in the future, the experiences using the BPI in clinical practice, and its functioning in a clinical setting will be explored. After all, little is known about using the BPI in clinical practice. Thus far, the results that have emerged from studies into the BPI are promising (Arseneault et al., 2005; Measelle et al., 1998; Ringoot et al., 2013), and suggest that the BPI can constitute a valuable supplement to youth care practices. When research from a clinical setting on use of the BPI is available, it may possibly be embedded in evidence-based practice (Mash & Hunsley, 2005). In conclusion, by means of this article we hope to have provided greater BPI publicity, to allow for optimal utilization of this instrument within youth care.
References


Bayer, J. K., & Sanson, A. V. (2000). Preventing the development of emotional mental health problems from early childhood: recent advances in the field. International Journal of Mental Health Promotion, 5, 4-16.


Abstract

Parental psychological control has been linked to symptoms of psychopathology in adolescence, yet less is known about its correlates in childhood. The current study is among the first to address whether psychological control is related to internalizing and externalizing problems in early childhood. A community sample of 298 children aged 7.04 (SD = 1.15) years participated. Along with two other parenting dimensions (i.e., responsiveness and behavioral control), psychological control, internalizing and externalizing problems were assessed by means of the Berkeley Puppet Interview. Psychological control was associated with internalizing and externalizing problems, and this association remained significant while controlling for parental behavioral control and responsiveness. Results suggest that the maladaptive correlates of psychological control also manifest in developmental periods prior to adolescence. Still, it is unknown how psychological control and child psychopathology are related over time in childhood.

Introduction

For quite a long time, research on parenting has taken a configurational approach, where parenting styles were defined as combinations of responsiveness (warmth) and demandingness (control) (e.g., Baumrind, 1971; Maccoby & Martin, 1983). Since the 1990s, parenting researchers increasingly adopted a dimensional approach, thereby studying the developmental course, correlates, and antecedents of separate parenting dimensions (e.g., Barber, 1996; Gray & Steinberg, 1999). Along with this shift to a dimensional approach came a more differentiated view on the concept of parental control. Specifically, a distinction was proposed between parental behavioral control and psychological control (Barber, 1996; Steinberg, 1990). Behavioral control was defined as parents’ regulation of children’s behavior through strategies such as limit-setting and monitoring. In contrast, psychological control was defined as control that intruded upon the child’s psychological world. Such control is characterized by manipulative and pressuring tactics including guilt induction, invalidation of the child’s perspective, and love withdrawal (Barber & Harmon, 2002). Because of its intrusive and pressuring nature, psychological control was hypothesized to represent a threat to children’s emerging sense of self and, as such, to increase the likelihood of maladjustment and internalizing problems in particular (Barber, 1996).

Initially, research on parental psychological control focused mainly on the developmental period of adolescence. There is robust evidence showing that psychological control is related to internalizing problems in adolescents (Barber & Harmon, 2002). This association has been shown (a) using parent and adolescent reports of psychological control (e.g., Petit, Laird, Dodge, Bates, & Criss, 2001; Soenens, Vansteenkiste, Duriez, & Goossens, 2006), (b) using cross-sectional and longitudinal research designs (e.g., Barber, Stolz, & Olsen, 2005; Soenens, Luyckx, Vansteenkiste, Duriez, & Goossens, 2008; Wang, Pomerantz, & Chen, 2007), and (c) to remain significant when controlling for other parenting dimensions, e.g., responsiveness and behavioral control (e.g., Barber, 1996; Gray & Steinberg, 1999; Soenens, Vansteenkiste, Luyten, Duriez, & Goossens, 2005).

Psychologically controlling parenting during early childhood

Research on parental psychological control in early childhood is comparatively scarcer. We think there are at least two reasons why the concept of psychological control has primarily attracted the attention of scholars interested in adolescence. First, internalizing distress, which represents the primary outcome of psychological control, increases during adolescence and represents a common threat to mental health during this developmental phase (Angold, Erkani, Silberg, Eaves, & Costello, 2002). We argue, however, that, although internalizing problems peak in adolescence, understanding its risk factors in early childhood may advance our understanding of
the development of internalizing problems (Cicchetti & Cohen, 2006; Luby, 2010). Second, psychological control has sometimes been portrayed as being independence-stilling in nature. For instance, Barber et al. (1994, p. 1121) defined psychological control as "as patterns of family interaction that intrude upon or impede the child’s individuation process, or the relative degree of psychological distance a child experiences from his or her parents and family.” That is, psychological control would hinder the development of a more independent orientation towards parents and a self-reliant stance. Given that processes of individuation and the development towards independence are highly salient during adolescence (Steinberg, 1990), one might expect that the detrimental effects of psychological control might be most pronounced in adolescence.

In later writings, Barber and colleagues (e.g., Barber & Harmon, 2002; Barber et al., 2005), emphasized that psychological control stifles not only independence but has a broader and more fundamental detrimental effect on children’s development of a stable and secure sense of self. On the basis of this later formulation one might argue that the effects of psychological control do not remain limited to adolescence but may also manifest in earlier and later developmental periods. Somewhat consistent with this argument, other scholars have argued that psychological control does not necessarily stifle independence but, instead, frustrates a more basic and universal psychological need for autonomy, that is, the need to feel a sense of volition and psychological freedom (Soenens & Vansteenkiste, 2010). According to self-determination theory (SDT; Deci & Ryan, 2000), this need for autonomy is universal and is not bound by age. Accordingly, parental psychological control would have deleterious effects on child adjustment irrespective of child age (Soenens & Vansteenkiste, 2010). Indeed, effects of parental psychological control have been reported for elementary-school aged children (i.e., 6-12-year-olds) regarding relational and physical aggression (Casas et al., 2006; Joussemet al., 2008; Kuppens, Grietens, Onghena, & Michielis, 2009a; Kuppens, Grietens, Onghena, & Michielis, 2009b), emotional distress (Nelson & Coyne, 2009; ) and to a lesser degree to internalizing problems (El-Sheikh, Hinnant, Kelly, & Erath, 2010). In toddlers, effects of psychological control on externalizing problems have been reported (Verhoeven, Junger, Van Aken, Dekovic, & Van Aken, 2010a), and psychologically controlling mothers have been found to have less compliant children in 6.5 year-olds (Verschueren, Dossche, Marcoen, Mahieu, & Bakermans-Kranenburg, 2006).

Given the existence of contrasting perspectives on the question whether psychological control is related to maladjustment across different ages, it was deemed important to examine whether parental psychological control is associated with internalizing and externalizing problems in early childhood. The current study addresses this issue by testing the association of child-reported psychological control with child-reported problems in 4-8 year-olds. In examining this association, we control either for externalizing or for internalizing problems, depending on which of these problems is used as the dependent variable, as comorbid problems are common in 4-8 year-olds (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003).

Although research on parental psychological control in 4-8 year-olds is scarce, recently research has begun to study the effects of parental psychological control on childhood problem behaviors (Aunola & Nurmi, 2005; Mills & Rubin, 1998; Olsen, Yang, Hart, Robinson, Wu, Nelson, et al., 2002; Rubin, Burgess, & Hastings, 2002; Van der Bruggen, Stams, Bögels, & Paulussen-Hoogeboom, 2010), with at least one study assessing psychological control with a measure that is sensitive to the developmental level of the child (Morris, Silk, Steinberg, Sessa, Avenevoli, & Essex, 2002). As yet, these studies have revealed rather inconsistent results. Morris et al. (2002) were one of the first to use age-appropriate techniques to measure child perceptions of psychological control by using a puppet interview to predict teacher-reported child adjustment. Morris et al. did not find direct effects, as the effects of psychological control were moderated by temperament.

In studies wherein different methods than the puppet interview were employed, results were also somewhat inconsistent. Psychological control was only linked to internalizing problems in 6-7-year-olds for children high on irritability. Aunola and Nurmi (2005) reported that high levels of parent-reported psychological control combined with high levels of parental affection affected child-reports of internalizing problems in children aged 5-6. Olsen et al. (2002) also found inconsistent evidence for a link between parent-reported psychological control and teacher-rated internalizing problems, in that this relation in 5 year-olds was moderated by gender and culture. Psychological control was related positively to internalizing problems for U.S. and Chinese girls, and Russian boys, but not for U.S. and Chinese boys or Russian girls.

As for observational studies focusing on psychological control, again, mixed results have been reported. Mills and Rubin (1998) showed that mothers of withdrawn-internalizing children aged 5-9 exhibited more psychological control during observed interactions than mothers of aggressive-externalizing children or children without problem behaviors. Rubin, Burgess, and Hastings (2002) found that, although observed maternal psychological control in 4-year-olds moderated the association of peer inhibition and social reticence, there was no direct association between maternal psychological control and internalizing problems. Finally, observed psychological control in children aged 4.5 on average has been found to correlate moderately with depression-anxiety, but this association was absent when controlling for children’s negative emotionality (Van der Bruggen et al., 2010). Given the strong associations of psychological control with internalizing problems in adolescence (Barber et al., 2005; Soenens & Vansteenkiste, 2010), it is remarkable that studies focusing on childhood have yielded much more inconsistent evidence.
Limitations of extant research

Extant research on psychological control and its association with internalizing problems in childhood suffers from at least two limitations. The first limitation is that in the majority of studies parent reports and observations of psychological control, internalizing, and externalizing problems were used. Parent reports of psychological control, as it represents a clearly maladaptive feature of parenting style, are particularly likely to be affected by social desirability (Bornstein, Hahn, & Haynes, 2011). To the best of our knowledge, only one study has examined whether young children are themselves able to report on parental psychological control (Morris et al., 2002). We assume that children are reliable informants of psychological control because child-reports of other parenting dimensions have been found reliable and valid in childhood (Sessa, Avenevoli, Steinberg, & Morris, 2001). A prerequisite for obtaining reliable and valid child-reports is to use developmentally appropriate instruments, such as the Berkeley Puppet Interview (BPI; Measelle, Ablow, Cowan, & Cowan, 1998). Furthermore, symptoms of internalizing problems in childhood such as guilt, worry, sadness and anhedonia, are, by definition, intra-individual experiences (Luby, 2010). Consequently, these symptoms are hard to detect for adult informants, leading to low inter-informant agreement of internalizing problems (Achenbach, 1987; De Los Reyes & Kazdin, 2005). Moreover, even for the observable behaviors often associated with externalizing problems, inter-informant agreement is low (Achenbach, 1987). Using child-reports may thus be key in assessing associations of psychological control with internalizing and externalizing problems.

The second problem is that psychological control has not often been examined in conjunction with other parenting dimensions, such as behavioral control and responsiveness (Barber, 1996; Soenens & Vansteenkiste, 2010). As a consequence, it has not been possible to examine whether the association between psychological control and childhood internalizing problems, if any, is unique (i.e., whether it remains significant after controlling for effects of the other parenting dimensions). Also, few studies have examined the possibility of interactions between psychological control and these other parenting dimensions in the prediction of children’s adjustment. This is unfortunate because at least some studies suggest that such interactions may occur. It has been shown, for instance, that the negative effects of maternal psychological control on both academic achievement (Aunola & Nurmi, 2004) and internalizing and externalizing problems (Aunola & Nurmi, 2005), are more pronounced when combined with high maternal responsiveness. Also, it has been suggested that a combination of behavioral control and psychological control may be particularly strongly related to externalizing problems (e.g., Pettit et al., 2001).

Study objectives

The present study sought to answer the following research questions. First, is child-reported parental psychological control related to child-reported internalizing problems, above and beyond behavioral control and responsiveness, and when controlling for externalizing problems? Second, is child-reported parental psychological control related to child-reported externalizing problems, above and beyond behavioral control and responsiveness, and controlling for internalizing problems. Third are the associations of psychological control with internalizing and externalizing problems moderated by age and gender? If, as can be argued on the basis of self-determination theory (Soenens & Vansteenkiste, 2010), psychological control frustrates a basic and age-invariant need for autonomy, one might expect that this association will hold across gender and age.

Method

Participants and Procedures

Parents of 1,575 children enrolled in the Kind in Zicht study, a cohort study about child mental health, were informed about this study by mail. Of the 480 children with active parental consent, 300 children were selected randomly to participate (50 % boys). Of these children, one was excluded because s/he was older than 8 years and one was excluded because data were missing, leading to a final sample of 298 children. Children’s age ranged from 5 to 8 years, with a mean age of 7.00 (SD = 1.13). The majority of children was of Dutch origin (97.4 %) and came from two-parent families (92.2 %). Of the children, 55.7 %, 36.8 % and 7.5% of their mothers had a high, medium and low educational level, respectively. Interviews were held at schools from January-February 2011. The interviews were conducted in a private setting, ensuring confidentiality, and were videotaped. When the interview was completed children received a package of stickers.

Measures

The Berkeley Puppet Interview (BPI; Measelle et al., 1998) is an interactive age-appropriate interviewing technique eliciting self-perceptions from 3.5-8 year-olds. The BPI is a reliable and valid instrument to assess child psychopathology and parenting (Measelle et al., 1998; Morris et al., 2002). BPI psychopathology scores show adequate temporal stability (r .56-.58, for externalizing and internalizing problems) and are correlated to measures of parent- and teacher rated psychopathology (r .30-.40, respectively) (Measelle et al., 1998). Furthermore, internalizing problems are meaningfully related to independent math and reading achievement test scores (r -.34, -.22, respectively) and externalizing problems to IQ, economic disadvantage.
and hyperactivity (r = -0.25, -1.17, 0.18, respectively) (Arsenault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005; Measelle et al., 1998). During the BPI children are interviewed by using two identical dog hand puppets, named Iggy and Ziggy. Before the interview starts, the puppets introduce themselves and explain the interview in a playful way. By using three practice items the interviewer judges whether the procedure is clear to the child and then proceeds to the interview or repeats the practice items until the procedure is clear. An example of a practice item is: Puppet 1: ‘I like chocolate’, Puppet 2: ‘I don’t like chocolate. How about you?’. Throughout the interview the puppets thus make opposing statements about themselves and then ask the child ‘How about you?’? The puppet with whom the child agrees then repeats the child’s answer, thereby confirming the child’s answer.

Interviews were conducted by three master students and one PhD-student. Interviewers were certified BPI administrators as they were trained during a two-day workshop and successfully completed eight practice interviews. Nuance was given to the BPI scores as interviews were coded by four trained observers on a 7-point scale. Responses that reflect the absence of psychopathology or particular parenting behaviors are coded 5, 6 or 7, depending on the weight the child puts in its answer. Whereas a 7 would reflect the highest end of presence (e.g., I’m a sad kid), the 6 would reflect the neutral absence response and the 5 a resistant response (e.g., Most of the time, I’m not a sad kid). At the opposite end of the spectrum, 1, 2 or 3, reflect presence of psychopathology or presence of particular parent behaviors. Here 1 refers to the highest end of presence (e.g., I’m always a sad kid), 2 reflects the neutral presence response and 3 again reflects a resistant response indicating presence of psychopathology (e.g., Most of the time, I’m a sad kid). When a child is not able to choose either one of the statements, this is coded a 4. To test whether coders were reliable, 15% of the videos were double-scored. Inter-rater agreement was satisfactory (Intraclass Correlation Coefficient; ICC .83).

Internalizing Problems were measured by using the 7-item Depression, 6-item Separation Anxiety and 7-item Overanxious BPI scales, resulting in a 20-item Internalizing Problems scale. Example items are: ‘I’m a happy kid/I’m not a happy kid’, ‘When I’m at school I don’t miss my mom/When I’m at school I miss my mom’, and ‘I don’t get bad dreams/I get bad dreams’.

Externalizing Problems were measured by using the 6-item Oppositional Defiant Behavior, the 9-item Conduct Problems and the 6-item Overt Hostility BPI scales, resulting in a 21-item Externalizing Problems scale. Example items are: ‘When I don’t get my way, I don’t yell at my teacher/When I don’t get my way, I yell at my teacher’, ‘I don’t tell lies/I tell lies’ and ‘I don’t hit kids/I often hit kids’.

Child perceptions of psychological control, behavioral control, and responsiveness were measured using adapted versions of the ‘Psychological Control Scale – Youth Self-Report (PCS-YSR; Barber, 1996), ‘Parental Regulation Scale’ – Youth Self-Report (PRS-YSR, Barber, 2002), and the Acceptance-Rejection scale from the Child Report of Parent Behaviors Scale (CRPBI; Schaefer, 1965), respectively. A validated Dutch translation of each of these scales was available (Soenens et al., 2006). A panel of four people independently adapted the items to the children’s developmental level and to match the BPI format. Subsequently, the adapted items were discussed in an expert panel and the final items were selected through consensus. The parenting dimensions were scored as described above for the BPI items.

Psychological control was measured with 8 items. Example items are: ‘When I upset my mom she will talk to me like she normally does/When I upset my mom she will only talk to me when I’m nice to her’, ‘When I’m telling a story, my mom does not interrupt me/When I’m telling a story, my mom interrupts me’. Behavioral control was measured with 8 items. An example item is: ‘My mom thinks it’s important that I keep to the rules/My mom doesn’t think it’s important that I keep to the rules’. Responsiveness was measured with 7 items. An example item is: ‘My mom comforts me when I’m sad/My mom doesn’t comfort me when I’m sad’.

Strategy for Analysis
As the BPI uses a bipolar answering method (i.e., agreeing with one puppet or the other to a certain degree), responses follow a bimodal rather than a normal distribution. Consequently, the frequency distribution of the 7-point BPI scores looks like a skewed normal curve with two peaks; One peak with variance around its center for one puppet, and one peak with variance around its center for the other puppet. The BPI is thus actually an index scale, not a Likert scale. In order to treat it as such, we developed a scoring system wherein we transformed BPI responses such that they resulted in a normally distributed scale. Each item was transformed into two new items; one positive and one negative item. This was computed as follows: the negative scores of the 7-point scale (i.e. the 1, 2, and 3 responses) and the neutral score were coded zero for the positive item, whereas the opposite procedure was followed for the negative item. For the positive item 5, 6 and 7 responses were recoded into, respectively, 1, 2 and 3 and for the negative item 1, 2 and 3 were recoded into 3, 2 and 1, hereby giving the most extreme score the highest weight. Subsequently, the sum score of the negative items were subtracted from the sum score of the positive items, thereby creating a difference score. These difference scores were used in subsequent analyses. An exception was made for the descriptive results, where mean scores (i.e., the mean score of the items with the 7-point scale) were computed in order to ensure comprehension and clarity of the constructs.

First, Pearson’s correlations were computed between all study variables. Second, six structural models were tested using MPlus. In the first model, internalizing problems were regressed on age, gender (coded: 1=boy; 2=girl), and psychological control. In the second model, behavioral control, responsiveness, and externalizing
problems were added to the model as independent variables. In a third model, interaction effects were tested between the various independent variables. These three models were also tested with externalizing problems as the dependent variable, and controlling for internalizing problems. For these analyses we used maximum likelihood estimation with the Huber-White covariance adjustment (MLR in Mplus 5.1; Muthén & Muthén, 1998-2007). Full Information Maximum Likelihood was employed such that all available data is used (FIML in Mplus 5.1; Muthén & Muthén, 1998-2007). The models we tested were saturated and had perfect fit ($X^2(0) = 0.00$).

Since children from the same classes may share common behaviors (i.e., clustering), ICC’s were calculated to determine the effects of class clustering. The ICC for internalizing problems was .004 and .00 for externalizing problems, indicating that only a very small part of the variance (.4 % and 0 %, respectively) could be explained by a class effect. Therefore, we decided to run the analyses without adjusting for clustering.

Results

Descriptive Statistics

Table 1 shows mean scores of children on internalizing and externalizing problems, psychological control, behavioral control, and responsiveness. As one might expect in a community sample, children experienced relatively low levels of internalizing and externalizing problems and psychological control, and relatively high levels of behavioral control and responsiveness. Furthermore, correlations between study variables are presented in Table 1. Psychological control was positively related to both internalizing and externalizing problems. Behavioral control was not related to internalizing problems, but was negatively related to externalizing problems. Responsiveness was negatively associated with internalizing and externalizing problems. Psychological control was not related to behavioral control and negatively related to responsiveness, while behavioral control was positively related to responsiveness. Age was negatively related to internalizing problems and positively to behavioral control and responsiveness. Gender was negatively related to externalizing problems and positively to psychological control, behavioral control, and responsiveness.
Primary Analyses

The first structural model tested the association between psychological control and internalizing problems. The model was saturated and standardized estimates are presented in Table 2. Results show that higher psychological control was associated with higher internalizing problems, while controlling for gender and age. Older children reported less internalizing problems than younger children did, and girls more than boys. In the second structural model we tested the association between psychological control and internalizing problems while controlling for behavioral control, responsiveness, and externalizing problems. This model was also saturated. Results show that psychological control remained uniquely associated with internalizing problems, while controlling for behavioral control and responsiveness. Externalizing problems were positively associated with internalizing problems. Girls reported more internalizing problems than boys, and older children reported less internalizing problems than younger children. Subsequently, in the third model, we tested interactions of psychological control with gender, age, behavioral control, and responsiveness, but none were significant.

In the second set of analyses, first the association between psychological control and externalizing problems was tested. The model was saturated and standardized estimates are presented in Table 3. Results show that higher psychological control was associated with higher externalizing problems, while controlling for gender and age. Boys reported more externalizing problems than girls did. In the second structural model we tested the association between psychological control and externalizing problems while controlling behavioral control, responsiveness, and for internalizing problems. This model was also saturated. Results show that psychological control remained associated with externalizing problems, while controlling for behavioral control and responsiveness. Behavioral control and responsiveness were negatively related to externalizing problems, while internalizing problems were positively associated with externalizing problems. Boys reported more externalizing problems than girls, and older children reported more externalizing problems than younger children. Subsequently, in the third model, we tested interactions of psychological control with gender, age, behavioral control, and responsiveness, but none were significant.

---

4 We also tested whether different results were obtained regarding the subscales of the internalizing scale; separation-anxiety, generalized anxiety, and depressive symptoms. Effects of psychological control on these subscales were all significant and in strength comparable to the reported effects of psychological control on the internalizing scale. No interaction effects were found. Please contact the first author for the detailed results of these analyses.

5 We also explored the role of age by testing for moderation of age group in multi-group analyses. We did not find any differences between these models, suggesting that the relationship between psychological control and problem behavior is not different for younger or older children in our sample. We decided not to report the results of these analyses here a) as the additional analyses did not yield any new results and b) for brevity reasons.

---

Table 2: Model 1 Predicting Internalizing Problems from Psychological Control, Model 2 Predicting Internalizing Problems from Psychological Control while Controlling for Parenting Dimensions and Externalizing Problems

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>Age</td>
<td>-2.95</td>
<td>.06</td>
</tr>
<tr>
<td>Gender</td>
<td>2.29</td>
<td>.06</td>
</tr>
<tr>
<td>Psychological Control</td>
<td>4.95</td>
<td>.05</td>
</tr>
<tr>
<td>Behavioral Control</td>
<td>-1.03</td>
<td>.06</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>0.07</td>
<td>.06</td>
</tr>
<tr>
<td>Externalizing Problems</td>
<td>0.36</td>
<td>.06</td>
</tr>
<tr>
<td>Psychological Control x Age</td>
<td>-0.16</td>
<td>.36</td>
</tr>
<tr>
<td>Psychological Control x Gender</td>
<td>-1.99</td>
<td>.19**</td>
</tr>
<tr>
<td>Psychological Control x Behavioral Control</td>
<td>-0.53</td>
<td>.09</td>
</tr>
<tr>
<td>Psychological Control x Responsiveness</td>
<td>-1.19</td>
<td>.13</td>
</tr>
<tr>
<td>Psychological Control x Externalizing Problems</td>
<td>-0.24</td>
<td>.18</td>
</tr>
</tbody>
</table>

Note. **p < .01, *p < .05, R² model 1 = .11**, R² model 2 = .15**. Gender coded 1=boy, 2=girl.
Discussion

The current study investigated the relation between child-reported psychological control and internalizing and externalizing problems in a sample of 4-8 year-olds. Results showed that psychological control was associated with internalizing problems, even when controlling for behavioral control, responsiveness and externalizing problems. Psychological control was also related to externalizing problems, and this relation remained when controlling for behavioral control, responsiveness and internalizing problems. Age, gender, behavioral control and responsiveness did not moderate the association of psychological control with internalizing problems or externalizing problems. From these results, we might conclude that parental psychological control has a robust association with both the internalizing and externalizing problem clusters, already in 4-8 year-olds. To our knowledge, this study is among the first to report a direct relation between child-reported psychological control and internalizing and externalizing problems in children. Whereas clear-cut evidence has been reported for a direct link between psychological control and internalizing problems during adolescence (e.g., Barber, 1996; Barber et al., 2005; Gray & Steinberg, 1999; Pettit et al., 2001; Soenens et al., 2005; Soenens et al., 2006; Soenens et al., 2008), evidence for such a direct link in childhood has not been reported previously. In addition, this study provides evidence that psychological control is also directly related to externalizing problems in early childhood, which is in line with studies conducted in adolescence (Barber & Olsen, 1997; Conger, Conger, & Scaramella, 1997; Eccles, Early, Frasier, Belansky, & McCarthy, 1997; Herman, Dornbush, Herron, & Herting, 1997; Mason, Cauce, Gonzales, & Higara, 1996).

Second, the contribution of psychological control was compared to two other important parenting dimensions, behavioral control and responsiveness. Psychological control showed up as a unique predictor of internalizing problems, which is in line with previous research (e.g., Barber, 1996), but nevertheless striking, as parental behavioral control and responsiveness are thought of as major correlates of child adjustment (Ainsworth, Blehar, Waters, & Wall, 1978; Hastings, Sullivan, McShane, Coplan, Utendale, & Vyncze, 2008; Patterson & Stouthamer-Loeber, 1984). Psychological control was not a unique predictor of externalizing problems as associations with behavioral control and responsiveness were also found. However, the effect of psychological control remained significant when including these parenting constructs to the model and was the strongest predictor of externalizing problems. These results point to a robust association of psychological control with externalizing problems and question whether internalizing problems should be seen as the primary outcome of psychological control, at least during early childhood.

Importantly, although psychological control was modeled as a predictor of internalizing and externalizing problems in the current study, the opposite direction of

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Model 1 Predicting Externalizing Problems from Psychological Control, Model 2 Predicting Externalizing Problems from Parenting Dimensions and Internalizing Problems</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE</td>
</tr>
<tr>
<td>Age</td>
<td>.135</td>
<td>.05</td>
</tr>
<tr>
<td>Gender</td>
<td>-3.68</td>
<td>.10</td>
</tr>
<tr>
<td>Psychological Control</td>
<td>7.40</td>
<td>.01</td>
</tr>
<tr>
<td>Behavioral Control</td>
<td>-2.02</td>
<td>.06</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>-2.25</td>
<td>.06</td>
</tr>
<tr>
<td>Internalizing Problems</td>
<td>-2.25</td>
<td>.06</td>
</tr>
<tr>
<td>Psychological Control x Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological Control x Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological Control x Behavioral Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological Control x Responsiveness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological Control x Internalizing Problems</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01, ***p < .001. Psychological Control and Internalizing Problems coded 1=boy, 2=girl.
effects also needs to be considered. Internalizing and externalizing problems may be predictive of parental psychological control (i.e., child effects model), and these factors may also mutually exert influence on each other (e.g., Belsky, 1984). Child effects and bidirectional associations indeed have been reported, although generally effects of adjustment problems on psychological control do not seem to be as strong as the effects of psychologically controlling parenting on maladjustment (Albrecht, Galambos, & Jansson, 2007; Kuppers et al., 2009a; Rogers, Buchanan, & Winchell, 2003; Soenens et al., 2008), with one exception in toddlers regarding externalizing problems (Verhoeven, Junger, Van Aken, Dekovic, & Van Aken, 2010b). Still, longitudinal research should test whether psychological control is predictive of internalizing problems, or vice versa, in these young children specifically.

**Working mechanisms of psychological control in its relation to problem behaviors**

In our view, the current findings provide some indirect insights into the nature of psychological control and the dynamics involved in its association with child adjustment. Sometimes it has been argued that psychological control has detrimental effects on children’s well-being and adjustment because it interferes with the development of an independent orientation. As an important developmental task in adolescence is to strive towards independence, and as psychological control would interfere with this developmental task, one may argue that psychological control has the strongest effects on internalizing problems during adolescence in particular. Continuing this line of reasoning, one might expect that psychological control is less strongly related or even unrelated to internalizing and externalizing problems in 4-8 year olds, when strivings for separation and independence are less salient compared to adolescence. The findings of this study contradict this reasoning because the strength of the association between psychological control and internalizing and externalizing problems seems to be comparable to the strength of associations obtained in adolescent samples (e.g., Barber, 1996; Barber et al., 2005; Barber & Olsen, 1997; Conger et al., 1997; Eccles et al., 1997; Gray & Steinberg, 1999; Herman et al., 1997; Petitt et al., 2001; Soenens et al., 2005). Hence, our findings suggest that parental stifling of independence may not be the working mechanism through which psychological control affects children’s internalizing and externalizing problems. In this regard, these findings mesh with recent research suggesting that psychological control does not necessarily stifle children’s independence (Kins, Soenens, & Beyers, 2012; Soenens, Vansteenkiste, & Sierens, 2009). Instead, it was shown in previous research that parents can use psychological control both to encourage dependence and to encourage independence. That is, through psychological control parents may pressure their children either to remain in close physical and emotional proximity (i.e., pressured encouragement of dependence) or to stand on their own two feet and make decisions without parental input (i.e., pressured encouragement of independence).

If psychological control is largely orthogonal to parental promotion of independence (versus dependence), then what is its working mechanism? Barber and colleagues (Barber & Harmon, 2002; Barber et al., 2005) have argued that psychological control is detrimental to the development of a secure and stable sense of self. Given that the formation of a positive and secure sense of self is a lifelong developmental task, it follows that psychological control might be harmful at all ages. Somewhat consistent with this argument, on the basis of self-determination theory, it has been argued that parental psychological control frustrates a basic and universal need for autonomy (Deci & Ryan, 2000; Soenens & Vansteenkiste, 2010). The need for autonomy is not defined in terms of independence but instead refers to feelings of initiative, volition, and self-endorsement that would be vital for all individuals’ thriving throughout the lifespan (Deci & Ryan, 2000). If indeed psychological control frustrates this basic psychological need for autonomy, it should impact development negatively irrespective of age and gender (Soenens & Vansteenkiste, 2010). Although it should be noted that the central tenets of self-determination theory, in particular the notion that psychologically controlling parenting would frustrate the basic psychological need for autonomy, were not tested directly in this study, we believe the current findings are generally consistent with the SDT-based reasoning. Associations between psychological control and internalizing and externalizing problems were already evident in young children and were not moderated by age and gender.

**Age specific aspects of psychologically controlling parenting on internalizing and externalizing problems**

Importantly, the notion that psychological control frustrates the need for autonomy at any given age does not imply that the specific way in which this autonomy frustration manifests is also age-invariant. Specifically, Soenens and Vansteenkiste (2010) have argued that the specific expression of autonomy frustration during a developmental period may depend on the psychosocial tasks that are most salient during that period. According to the psychosocial stages model, children aged 4-8 face the developmental task of developing a sense of initiative (Erikson, 1968). During this stage, children take initiative to increasingly investigate and explore the physical and social world. When this developmental task is frustrated, children may develop a reluctance to explore as curiosity arouses a sense of guilt (Newman & Newman, 2009). When psychologically controlling parenting thwarts the resolution of this stage-salient psychosocial task, it is likely to be expressed in terms of a lack of initiative and proneness to debilitating feelings of guilt that, in turn, increase the likelihood of internalizing problems (Soenens & Vansteenkiste, 2010). Related to this hypothesized mechanism, psychological control has been hypothesized to affect
internalizing problems by reducing perceived personal control (Weisz, Southam-Gerow, & McCarty, 2003), reducing perceived mastery (Chorpita & Barlow, 1998), and inducing perceived helplessness (Garber & Flynn, 2001). Furthermore, early childhood is characterized by compliance issues, possibly in interaction with growing initiative attempts of the child. When parents react in a psychologically controlling fashion to these compliance issues, this might result in noncompliance, resistance, and other externalizing behaviors (Patterson, Reid, & Dishion, 1992). Also, frustration of the need for autonomy has been hypothesized to affect externalizing behaviors by impeding the internalization process (Ryan, Deci, Grolnick, & La Guardia, 2006), thwarting self-regulation capacities and inducing negative affect (e.g., Assor, Roth, & Deci, 2004), which are thought of as key psychological factors in the development of aggression (e.g., Eisenberg, Fabes, Nyman, Bernzweig, & Pinuelas, 1994). Clearly, more research is needed to examine mediating mechanisms of parental psychological control in young children.

Limitations, future directions, and implications

A number of limitations should be noted. First, all relations were studied cross-sectionally and thus, are of a correlational nature. Therefore, these results do not evidence that psychological control predicts internalizing problems and do not provide insight into its developmental course. Future research should study these relations longitudinally to address the question whether psychological control influences the development of internalizing problems or whether children who show more internalizing problems evoke more psychological control. A second limitation is that one informant was used for the assessment of all variables. We relied on child reports of psychological control because it has been argued that the subjective experience of psychological control, rather than its objective display, is ultimately the strongest determinant of child adjustment (Barber, 1996). More generally, it has been argued that the perception of parenting practices may be more important than actual parenting when examining child adjustment (Fuligni & Eccles, 1993). This being said, we strongly recommend the use of multiple informants in future research because such an approach would allow one to avoid problems associated with shared method variance (De Los Reyes & Kazdin, 2005). It would be particularly interesting to include both objective measures of psychological control (e.g., ratings of observed parent-child interactions; Barber, 1996) and measures of subjectively experienced psychological control. By including both types of measures, it will be possible to determine the degree of convergence between objective displays of psychological control and subjective experiences thereof as well as to model the shared variance between both types of measures as a more valid assessment of psychological control.

In addition, such a study would allow one to address important content-based questions related to how children perceive parental behavior. One may wonder, for instance, why some children are prone to experience even small and infrequent displays of psychological control as strongly intrusive while other children are relatively less likely to experience objective parental behavior as psychologically controlling. Possibly, children’s temperament and personality play an important role. Children scoring higher on negative reactivity might be more likely to experience parental behavior as intrusive and, in addition, may develop more internalized and externalized distress in response to such experiences. In sum, including both objective and subjective assessments of psychological control would allow for new ways of examining interactions between parental behavior and children’s individual differences in predicting important developmental outcomes.

Relatedly, although only direct relations of psychological control with internalizing and externalizing problems were found, we did not include measures of temperament or culture. As former studies did find interactions between either temperament and psychological control and culture and psychological control (Mills & Rubin, 1998; Morris et al., 2002; Olsen et al., 2002; Rubin et al., 2002; Van der Bruggen et al., 2010), future studies should attempt to investigate these interactions using a similar design as the current study. As such, it may be elucidated whether psychological control indeed has a main effect on adjustment problems, or that relations may be dependent on other constructs.

Another aim for future research might be to test more directly the SDT-based hypothesis that controlling parenting and subsequent need frustration may forestall the process of internalization and, through this effect, may lead to undesirable behavior (e.g., lack of compliance and disruptive behavior) (e.g., Grolnick, Deci, & Ryan, 1997). Indeed, in SDT it is assumed that children may follow parental rules and guidelines for relatively more internalized (i.e., autonomous) reasons (e.g., because they understand and identify with the importance of the rules) or for relatively less internalized (i.e., controlled) reasons (e.g., to avoid punishment or to avoid pressuring feelings of guilt and shame). Psychologically controlling parenting and externalizing behaviors are assumed to relate to less internalized reasons for adherence to parental rules. Low internalization may, in turn, predict less compliance and more disruptive behavior in the long term (Soenens & Vansteenkiste, 2010). Future research could test this hypothesized sequence of events, which already received some support in elementary school children and adolescents (e.g., Grolnick, Ryan, & Deci, 1991; Soenens, Vansteenkiste, & Niemiec, 2009), among preschool children.

Finally, although the current study used SDT as a framework, we did not test its tenets directly. Specifically, we did not actually test whether psychological control is related to frustration of the needs for autonomy, competence, and relatedness. In order to draw firmer conclusions about the working mechanism of psychological control...
control, future studies should include measures that tap into need satisfaction. Such evidence is available in research with adolescents (e.g., Ahmad, Vansteenkiste, & Soenens, 2013), yet not in research with preschool children. For the purpose of similar research with preschool children, items from existing questionnaires on need satisfaction (e.g., Gagné, 2003) could be adapted to the BPI-format.

Results from this study may have potential implications for both intervention and prevention. Often, parenting programs do not take psychoeducational content into account. For example, Beardslee and colleagues designed a prevention program that aims to reduce internalizing problems by increasing positive interactions and understanding of parental depression (Beardslee, Gladstone, Wright, & Cooper, 2003), and Triple P focuses on improving parenting skills by increasing parental use of effective consequences to decrease maladaptive behaviors (Sanders, Turner, & Markie-Dadds, 2002). These parenting skills mostly deal with parental provision of clear guidelines, consistency, and monitoring of the child’s behavior, aspects of parenting that could be referred to as behavioral control or structure.

Although this focus on structure is important, parenting programs may tend to neglect the way parents communicate these strategies to their children. We believe psychological control is an important feature of how parents communicate structure. For example, providing psycho-educational information about parental depression in a psychologically controlling versus autonomy-supportive style (Soenens & Vansteenkiste, 2010). However, solid implications for practice should be empirically supported by more research. Therefore, first and foremost, research into the longitudinal relations between parental psychological control and childhood psychopathology is warranted.

References


Barber, B. K. (2002). Regulation as a multicultural concept and construct for adolescent health and development. Unpublished manuscript.


Chapter 10

Relations Between Parental and Child Separation Anxiety:
The Role of Dependency-Oriented Psychological Control

Submitted for publication:
Stone, L. L., Otten, R., Soenens, B., Engels, R. C. M. E., & Janssens, J. M. A. M.
INTRODUCTION

Separation anxiety is a developmentally appropriate reaction of distress to separation of the caregiver during infancy, and central to the child’s psychological development (Blatt, 2004; Bowlby, 1988; Mahler, 2000). Although most children adequately learn to regulate their distress reaction to separation, some children continue to experience anxiety following separation. When symptoms of separation anxiety persist, these behaviors can become highly problematic and debilitating (Jurbergs & Ledley, 2005). In this case, normative fears and worries concerning separation of the caregiver become non-age-appropriate and are associated to school refusal and excessive truancy (Egger, Costello, & Angold, 2003; Kearney & Albano, 2004). Separation anxiety disorder is the most common anxiety classification below the age of 12 (Cartwright-Hatton, McNicol, & Doubleday, 2006). It is an antecedent of adult anxiety disorders and linked to depression in young adults (Hirschfeld-Becker, Micco, Simoes, & Henin, 2008). Symptoms of separation anxiety have been found to be more influence by the shared environment than by heritability, therefore, gaining insight into familial factors that may maintain or exacerbate separation anxiety seems to be very important (Eley, Rijsdijk, Perrin, O’Connor, & Bolton, 2008). The aim of the present study is to investigate whether maternal separation anxiety is related to separation anxiety in children aged 5-9, and, if any, to elucidate parenting processes that may explain these relations.

In developmental theories, great importance is attributed to issues of separation between mother and child, as it is central to the development of the child’s psychological self in a process termed separation-individuation (Mahler, 2000). Also, studying separation may be central to understanding aspects of parenting (Hock, McBride, & Gnezda, 1989). Object relations theory proposes that there is an optimal maternal distance, suited to the infant’s changing developing needs (Blatt, 2004; Mahler, 2000). It is suggested that when mothers do not have a healthy sense of self, it is more difficult to see herself as separate from the child, and experiences of separation bring about personal feelings of loss or rejection (Hock & Schirtzinger, 1992). In turn, these feelings may lead to worry, sadness, or guilt that color the separation experience. This process is termed maternal separation anxiety (Hock et al., 1989). The threatening feelings regarding separateness that are characteristic of maternal separation anxiety may lead to more protective behaviors in the mother (Barber & Harmon, 2002), and as such impede the separation-individuation process in children (Blatt, 2004). Although research has begun to study some of these theoretical propositions, the overall line of reasoning requires empirical testing. A few studies have investigated the link between maternal separation anxiety and separation anxiety in their children (Dallaire & Weinraub, 2005; Hock, Hart, Kang, & Lutz, 2004; Mayseless & Scher, 2000; Peleg, Halaby, & Waby, 2006). In a cross-sectional study,
maternal perception of separation effects on her child was related to observed child separation anxiety in 36 preschoolers (Peleg et al., 2006). Further, maternal separation anxiety predicted symptoms of separation anxiety in 99 children at age 6 (Dallaire & Weinraub, 2005), and prior maternal worry predicted feelings of anxiety in 48 11-year-olds (Hock et al., 2004). In contrast, maternal separation anxiety was not related to fearful temperament in 97 infants (Mayseless & Scher, 2000). These studies provide some evidence that maternal separation anxiety is related to child separation anxiety, but it remains unclear how these anxieties are transferred from mother to child.

One construct that may function as an explanatory mechanism of the link between maternal separation anxiety and child separation anxiety is psychologically controlling parenting. Psychological control is an intrusive parenting tactic, characterized by pressuring and manipulative strategies such as love withdrawal, guilt induction, and conditional approval (Barber, 1996). There is robust evidence showing that psychological control is related to internalized distress in adolescents (Barber, Stoltz, & Olsen, 2005; Soenens, Luyckx, Vansteenkiste, Duriez, & Goossens, 2008), and some in children (Stone et al., 2013a). In this study, we argue that one specific form of psychological control may explain the link between maternal and child separation anxiety. Theory and research suggest that maternal separation anxiety is associated with psychological control. Threatening feelings regarding increasing separateness are argued to lead to more restrictive parenting behavior, in order to keep the child in close proximity (Barber & Harmon, 2002; Soenens, Vansteenkiste, Duriez, & Goossens, 2006). In line with this reasoning, parental anxieties regarding separation were related to psychological control in adolescents (Soenens et al., 2006). Subsequently, Soenens, Vansteenkiste and Luyten (2010) proposed that maternal difficulties with interpersonal relatedness and closeness may lead to specific controlling parenting tactics, termed dependency oriented psychological control. It is argued that love and care are made contingent on the child’s dependence on the parents. Indeed, dependency-oriented psychological control is strongly related to parental anxiety regarding separation in adolescents (Soenens et al., 2010).

Regarding dependency-oriented psychological control and child separation anxiety, theory and research also suggest a link between these constructs. Psychological control is hypothesized to represent a threat to the child’s emerging sense of self (Barber, 1996), as the child may be unable to develop a stable representation of the mother as a caring person. This unstable representation of the mother may lead to fears of loss of love and abandonment when the child attempts to separate from the parent (Blatt, 2004), potentially leading to difficulties in distancing, interpersonal differentiation, and boundary-formation for the child (Hock & Schritzinger, 1992). In line with the notion that parenting tactics aimed at keeping the child in close proximity are associated with anxiety regarding separation, it was found that dependency-oriented psychological control was related to dependent personality features and depressive symptoms in adolescents (Soenens et al., 2010). Another proposed mechanism is that children of intrusive and controlling parents lack experience with independence and may perceive novel and ambiguous situations as threatening, which may provoke anxiety in separation-laden contexts (Wood, 2006). Accordingly, it has been found that intrusive parenting and separation anxiety are related in children (Wood, 2006), and adolescents (Kins et al., 2011; Mayseless & Scharf, 2009).

In sum, we argue that maternal psychological control may be an intervening variable between maternal separation anxiety and her child’s separation anxiety. To our best knowledge, only one study explicitly tested this proposed mechanism. In this study, the relation between maternal separation anxiety and separation-individuation pathology in a large sample of emerging adults was found to be partially mediated by dependency oriented psychological control (Kins, Soenens, & Beyers, 2011). However, no studies have tested this mechanism in childhood, where separation anxiety is most salient (Cartwright-Hatton et al., 2006). The present study sought to answer the following research questions. First, is maternal separation anxiety related to separation anxiety in children? Second, is this relationship mediated by dependency oriented psychological control? Third, are these relations moderated by gender? Based on previous studies, we expect that maternal separation anxiety is associated with child separation anxiety and partially mediated by dependency oriented psychological control. In accordance with the notion that cultural stereotypes render females more vulnerable to problems with interpersonal relatedness and dependency, we hypothesize that mothers use more dependency promoting techniques towards their daughters (Blatt, 2004), and that associations between maternal separation anxiety and their daughters’ separation anxiety are stronger than between mothers and sons.

**Method**

**Participants and Procedures**

In this study, 300 children were interviewed during the first measurement (T1). One child was excluded due to missing data and another child because she was over eight years old. One year later (T2), 288 of these children (96%) were re-interviewed, of whom one was excluded because of her advanced age. This resulted in a sample of 298 children at T1, and 287 children at T2. Of these participating children, 50% was male and the mean age was 6.95 years (SD=1.13; range 5-8 years). The majority of the children was of Dutch origin (97.4%) and grew up in a two-parent family (92.2%).

Parents (T1 n=289, T2 n=269) completed questionnaires about the children at both time points. The parents who filled out the questionnaires were on average 38.29 years old, 90% were female, and the mean age was 38.5 years (SD=5.8; range 25-57 years). The mean age of the children was of Dutch origin (97.4%) and grew up in a two-parent family (92.2%).
years old (SD = 3.88), and 92.9% of them were female. Over half of the parents were highly educated (54.8%), 37.3% had an intermediate education level, and 6.6% lower education. Slightly over 1% received some other type of education.

Longitudinal data (2011(T1)-2012(T2)) from the Child in Sight project were used (Stone et al., 2013b), which was approved by the committee on ethics from the Radboud University Nijmegen. Within this project, information was collected via children and their parents. Informed consent from the children's parents was obtained. Each year, the Berkeley Puppet Interview (BPI; Measelle et al., 1998) was administered to the children by five certified master students or researchers. They all completed a training course in which the interviewing techniques of the BPI were extensively practiced. Subsequently, they each conducted eight practice interviews, and were then evaluated. The interviews were administered at primary schools in January and February of 2011 and 2012. Children were interviewed in an empty classroom to ensure confidentiality. Interviews were videotaped and after completion, the children received a pair of stickers to thank them for their participation.

Measures

**Maternal Separation Anxiety.** At both waves mothers rated their feelings of separation anxiety concerning their child (MSAS; Hock et al., 1989). The subscale Maternal Separation Anxiety, which consists of 21 items was used for this study. Each item is answered on a 5-point Likert scale, ranging from strongly disagree (1) to strongly agree (5). Sample items include ‘My child is happier when s/he is with me than when s/he is with the babysitter or teacher’ and ‘When I’m not with my child, I don’t have fun’. Psychometric properties of the MSAS have been found adequate. Cronbach’s alpha ranged from .83-.84.

**Separation Anxiety.** The BPI is an interactive age-appropriate interviewing technique eliciting self-perceptions from 3.5-8-year-olds. The BPI has proven a reliable and valid instrument to assess child psychopathology and parenting (Measelle et al., 1998; Morris et al., 2002; Ringoot et al., 2013). BPI psychopathology scores are correlated to measures of parent- and teacher rated psychopathology (r .30-.40, respectively) (Measelle et al., 1998). During the actual BPI interviews children are interviewed by using two identical dog hand puppets, named Iggy and Ziggy. Before the interview starts, the puppets introduce themselves and explain the interview in a playful way. By using three practice items the interviewer judges whether the procedure is clear to the child and then proceeds to the interview or repeats the practice items until the procedure is clear. An example of a practice item is: ‘Puppet 1: ‘I like chocolate’, Puppet 2: ‘I don’t like chocolate. How about you?’: Throughout the interview the puppets thus make opposing statements about themselves and then ask the child ‘How about you?’ The puppet with whom the child agrees then repeats the child’s answer, thereby appraising the child’s answer.

**Dependency Oriented Psychological Control.** Child perceptions of dependency oriented psychological control were measured using adapted versions of the ‘Psychological Control’ Scale – Youth Self-Report (PCS-YSR; Barber, 1996), ‘Parental Regulation Scale’ – Youth Self-Report (PRS-YSR; Barber, 2002), and the Acceptance-Rejection scale from the Child Report of Parent Behaviors Scale (CRPBI; Schaefer, 1965), respectively. A validated Dutch translation of each of these scales was available (Soenens et al., 2006). A panel of four people independently adapted the items to the children’s developmental level and to match the BPI format. Subsequently, the adapted items were discussed in an expert panel and the final items were selected through consensus. The parenting dimensions were scored as described above for the BPI items. Dependency-oriented psychological control was measured with four items. Example items are: ‘When I upset my mom she will talk to me like she normally does/When I upset my mom she will only talk to me when I’m nice to her’, ‘When I’m telling a story, my mom does not interrupt me/When I’m telling a story, my mom interrupts me’.

**Strategy for Analysis**

First, Pearson’s correlations were computed between all study variables. Second, three path models were estimated in Mplus version 6 (Muthén & Muthén, 1998-2007) to evaluate the effects of maternal separation anxiety on children’s separation anxiety via dependency-oriented psychological control. The first model evaluated these associations cross-sectionally at T1. The second two models evaluated these
associations longitudinally, with the second model not controlling for T1 separation anxiety in children, and the third model controlling for T1 separation anxiety in children. Model fit was assessed with various fit indices, including robust chi-square with estimated degrees of freedom (df), comparative fit index (CFI; Bentler, 1990), root mean squared error of approximation (RMSEA; Byrne, 1998), and Tucker–Lewis index (TLI; Tucker & Lewis, 1973). Direct associations between variables were assessed based on standardized path coefficients and p-values. Indirect effects (i.e., mediation) were tested using a bootstrap method in Mplus (Shrout & Bolger, 2002). Missing values on predictor variables were substituted in Mplus using full information maximum likelihood (FIML) estimation.

To examine gender differences in individual model paths, a freely estimated model was compared to a model in which parameters were constrained to be equal for boys and girls. If a significantly worse fit to the data was found for the constrained model, we employed a stepwise approach, such that each of the parameters were tested separately for gender differences. A chi-square difference test was used to test relative model fit (Satorra & Bentler, 2001).

Results

Descriptive statistics
Maternal separation anxiety at T1 was strongly related to maternal separation anxiety at T2 (Table 1). Further, dependency-oriented psychological control at T1 was related to dependency-oriented psychological control at T2. A similar association was found for separation anxiety, indicating temporal stability of these problems in young children. Maternal separation anxiety at T1 was related to dependency-oriented psychological control and separation anxiety concurrently, but not longitudinally. Dependency-oriented psychological control was related to maternal separation anxiety and separation anxiety in children both concurrently and longitudinally. At both waves, more separation anxiety was reported by girls than boys, respectively (t(296) = -2.38, p < 0.05) and (t(285) = -2.82, p < 0.01). No gender differences were found for the reports of maternal separation anxiety and dependency-oriented psychological control. Finally, on average, children reported more separation anxiety at T1 than at T2 (t(284) = 4.18, p < 0.01), and more maternal separation anxiety was reported at T1 than at T2 (t(217) = 3.15, p < 0.01).

Cross-sectional mediation model
The cross-sectional mediation model had adequate fit ($\chi^2$(2) = 2.27, $p = .32$; CFI = .990; RMSEA = .023 (CI = .000–.128); TLI = .966). A positive trend was found between maternal separation anxiety and separation anxiety in children (beta = .13, SE = .90, $p = .05$). Dependency-oriented psychological control was positively related to separation anxiety in children (beta = .24, SE = .09, $p = .001$). Maternal separation anxiety and dependency-oriented psychological control were positively marginally related (beta = .13, SE = .65, $p = .059$). Further, age was negatively related to separation anxiety in children (beta = -.13, SE = .27, $p = .021$), such that younger children reported more separation anxiety than older children. Gender was not related to separation anxiety (beta = .10, SE = .65, $p = .10$). The association between maternal separation anxiety and child separation anxiety was not statistically mediated by dependency-oriented psychological control (indirect effect = .03, SE = .02, $p = .09$). The model explained 10.4% of the variance of separation anxiety in children.

Longitudinal mediation models
In the first longitudinal model, we did not control for separation anxiety at T1. This model had adequate fit ($\chi^2$(2) = 5.52, $p = .06$; CFI = .987; RMSEA = .022 (CI = .000–.127); TLI = .956). Maternal separation anxiety at T1 was not related to separation anxiety in children at T2 (beta = .07, SE = .96, $p = .24$). Dependency-oriented psychological control at T1 was not related to separation anxiety in children at T2 (beta = -.11, SE = .11, $p = .09$). Maternal separation anxiety at T1 was marginally related to dependency-oriented psychological control concurrently (beta = .13, SE = .64, $p = .056$). Further, age was again negatively related to separation anxiety at T2 (beta = -.19, SE = .31, $p = .001$), such that younger children reported more separation anxiety than older children. Gender was positively related to separation anxiety at T2 (beta = .17, SE = .74, $p = .007$), such that girls reported more separation anxiety than

<table>
<thead>
<tr>
<th>Table 1 Correlations between all study variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>M (SD)</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>1 Maternal Separation Anxiety T1</td>
</tr>
<tr>
<td>2 DPC T1</td>
</tr>
<tr>
<td>3 Separation Anxiety T1</td>
</tr>
<tr>
<td>4 Maternal Separation Anxiety T2</td>
</tr>
<tr>
<td>5 DPC T2</td>
</tr>
<tr>
<td>6 Separation Anxiety T2</td>
</tr>
<tr>
<td>7 Age T1</td>
</tr>
</tbody>
</table>

Note. DPC = Dependency-Oriented Psychological Control, ** p < .01, * p < .05
boys. The association between maternal separation anxiety at T1 and child separation anxiety at T2 was not statistically mediated by dependency-oriented psychological control (indirect effect = .02, SE = .01, p = .26). The model explained 7.7% of the variance of separation anxiety at T2 in children.

The model in which we controlled for separation anxiety at T1 showed adequate fit to the data ($\chi^2$ (1) = .001, $p = .97$, $CFI = 1.000$; RMSEA = .000 (CI = .000-.000); $TLI = 1.122$). Separation anxiety in children at T1 was associated with child separation anxiety at T2 (beta = .32, SE = .07, p = .000). Maternal separation anxiety at T1 was not related to separation anxiety in children at T2 (beta = .03, SE = .07, p = .57). Dependency-oriented psychological control at T1 was not related to separation anxiety in children at T2 (beta = .05, SE = .10, p = .44). Maternal separation anxiety at T1 was positively related to dependency-oriented psychological control concurrently (beta = 4.37, SE = .13.9, p = .003). Further, age was again negatively related to separation anxiety at T2 (beta = -1.6, SE = .28, p = .003), and gender was positively related to separation anxiety at T2 (beta = .14, SE = .63, p = .011). The association between maternal separation anxiety at T1 and child separation anxiety at T2 was not statistically mediated by dependency-oriented psychological control (indirect effect = .22, SE = .17, p = .20). The model explained 18% of the variance of separation anxiety at T2 in children.

Gender differences

Imposing constraints on the path loadings did not result in statistically significant model comparisons; the constrained models did not differ from the unconstrained models. These results indicate that the path coefficients do not differ among boys and girls.

Discussion

The current study investigated whether maternal separation anxiety and separation anxiety in children were related cross-sectionally and longitudinally in a sample of 5-9-year-olds using innovative and age-appropriate measures for child-reports. Second, we investigated whether this relation was mediated by dependency-oriented psychological control. Results showed that maternal separation anxiety was related to separation anxiety in children cross-sectionally, albeit weakly, and not longitudinally. Moreover, dependency-oriented psychological control was related to separation anxiety in children cross-sectionally, but not longitudinally. Maternal separation anxiety was marginally related to dependency-oriented control. Further, there was no mediation of dependency-oriented psychological control, nor cross-sectionally or longitudinally. In conclusion, our findings do not support our hypotheses.

Studies on the association between maternal and child separation anxiety are scarce, and the available studies are hampered by small sample sizes (Dallaire & Weinraub, 2005; Hock et al., 2004; Mayseless & Scher, 2000; Peleg et al., 2006). Moreover, these studies were conducted in different developmental periods, making it hard to compare their findings. Although these studies were based on theoretical assertions (Blatt, 2004; Hock et al., 1999), the theory is quite unspecific as to when maternal separation anxiety should impact children, and for whom this may be most disturbing. In the current study, we found that maternal separation anxiety was weakly related to separation anxiety in children, and we could not replicate previous findings regarding an association between maternal separation anxiety and separation anxiety in children over time. This shows that the available evidence for a link between maternal separation anxiety and separation in children is not supported. Although Dallaire and Weinraub (2005) found a quite strong association between maternal separation anxiety and child separation anxiety in children over time, in the same developmental period as the children in our study, these authors measured maternal separation anxiety during infancy. This may be important, as it may indicate that current feelings of maternal separation anxiety may be not as important as feelings of maternal separation anxiety earlier in development. This reasoning coincides with the theoretical proposition that the first process of separation-individuation is hypothesized to take place during infancy (Mahler, 2000), and with attachment research, where ample studies have shown that early mother-child interactions shape children’s ability to regulate their emotions (e.g., Bakearmans-Kranenburg, Van IJzendoorn, & Juffer, 2003).

Therefore, other factors may be more important predictors of separation anxiety at this age than maternal separation anxiety. Although it is unclear whether well-known parenting tactics, such as low responsiveness and behavioral control predict separation anxiety in children, maternal sensitivity has been reported as predicting separation anxiety (Dallaire & Weinraub, 2005). Further, scholars using a cognitive framework for studying separation anxiety have shown that separation-related interpretive biases in children are related to childhood separation anxiety (Bögels, Sniider, & Kindt, 2003; In-Albon, Dubi, Rapee, & Schneider, 2009; In-Albon, Kossowsky, & Schneider, 2010; Perez-Olivas, Stevenson, & Hadwin, 2011). However, these findings may be hard to replicate in more ecologically valid research, such as the current study. Also, these experimental studies did not test whether these biases predict separation anxiety in children over time.

Second, dependency-oriented psychological control did not mediate the relation between maternal separation anxiety and child separation anxiety. This contradicts the findings of Kins, Soenens and Beyers (2011), who found dependency-oriented psychological control partially mediated the relation between maternal separation and separation issues in early adults. This divergence in findings may be due to the...
developmental period at hand. Our study focuses on early childhood where the child is not yet expected to separate from the mother. The study of Kins et al. focuses on early adults, where disconnecting from the parent and becoming independent is a salient developmental task (Arnett, 2000). Possibly, younger children do not experience their mothers as being afraid of distancing, as these children are usually in close proximity to their mother. Dependency-oriented psychologically controlling parenting thus does not seem to be the mechanism whereby separation anxieties are transferred from mother to child during early childhood.

Third, we did not find that children experienced their mothers to use more dependency-oriented psychological control toward girls than boys. Further, associations between maternal separation anxiety and separation anxiety in children were not stronger for girls than for boys. Thus, our hypothesis regarding the moderating role of gender in separation anxiety was not supported. Therefore, the notion that females may be more vulnerable to problems with interpersonal relatedness and dependency due to cultural stereotyping (Blatt, 2004), was not confirmed by this study. As no previous studies tested this hypothesized association (Blatt, 2004), these findings primarily call for replication of studies using a similar design and sample.

A number of limitations to this study should be noted. First, our measure of dependency-oriented psychological control is rather short. This may impede the reliability and validity of this scale. Thus, it is questionable whether we really measured what we intended to measure. This being said, correlations show that there is stability of this construct, and that it is correlated to measures it is theoretically expected to be associated with. Still, future research should study this construct in adolescents, such that a well-validated self-report instrument can be employed (e.g. Soenens et al., 2010). Second, although we used multiple informants in this study, we did not include alternative measures of both our maternal and separation anxiety measures, and dependency-oriented psychological control. It has been argued that when measuring a construct, multiple methods should always be employed, as to be certain of the veracity of the measurement (De Los Reyes, 2013). For example, including different measures enables assessing the degree of convergence between the measures, thereby probably increasing validity of the construct. Such alternative methods could include questionnaires, but also observations of parent-child interactions (Barber, 1996). Third, our sample is biased in that roughly half of our participants are highly educated. This may have led to typical problems found with highly educated samples, namely that there is less variance in problem behaviors.

In conclusion, this study found no associations between maternal separation anxiety and separation anxiety in children over time in young children, while using a large sample and different informants. Also, we did not find support for the mediating role of dependency-oriented psychological control. Future research is warranted in order to draw firm conclusions about the relations between maternal and child separation anxiety, and the possible mediating role of dependency-oriented psychological control. For future research the recommendations mentioned above should be taking into account.
Chapter 11

General Discussion
Summary of the main findings

The aim of part I of this thesis was to examine the psychometric properties of the Strengths and Difficulties Questionnaire and to examine to what extent contextual factors are associated with childhood psychopathology during early childhood. The aim of part II of this thesis was to examine the psychometric properties of a child self-report instrument, the Berkeley Puppet Interview, and to investigate the role of parenting tactics in childhood psychopathology.

In this concluding chapter the results concerning these research questions will be discussed. As we discussed the issues pertaining to each study specifically in the previous chapters, we will focus our discussion here on the key overarching themes important to the current thesis. Next, we reflect on the limitations of this thesis, present possible directions for future research and describe the implications of our findings.

Part I - Problem behaviours in young children: psychometrics and contextual factors

- A review of studies investigating psychometric properties of the SDQ showed that reliability as indexed by Cronbach’s alpha of its subscales is not always adequate, particularly so for the parent version. Also, this review showed predictive validity is scarcely investigated.

- Regarding the parent version of the SDQ, we found adequate reliability indices by using an alternative indicator of reliability, namely coefficient Omega. The SDQ’s five-factor structure was replicated. Measurement invariance was established for gender, survey method, maternal age, educational level, and sibship size.

- Regarding the parent and teacher version of the SDQ, we found indicators of reliability and test-retest reliability to be adequate and indicators of construct validity, criterion validity, and predictive validity to be of good quality. We established measurement invariance for the parent version regarding gender, age and ethnicity. The most stringent type of measurement invariance was established for the teacher version regarding age and ethnicity, not for gender.

- In early childhood, parent reported externalizing and internalizing problems were not reciprocally related over time. However, when taking problem level into account, externalizing problems were related to clinical level internalizing problems over time, even when controlling for a selection of third variables.
• Young children are substantially similar to friends regarding teacher reported internalizing problems, when friends both indicate that they are friends (i.e., reciprocated friendship). Children are not similar to a peer they perceive as a friend, while this peer does not perceive the child as a friend (i.e., unilateral friendship).

• Parent reported internalizing problems remain stable during early childhood, while externalizing problems decrease during this developmental phase. Parenting stress, as reported by parents, also decreases during early childhood. Interrelated development between internalizing and externalizing problems on the one hand and parenting stress on the other was found.

Part II - Child perceptions of problem behaviours and parenting

• Stability and criterion validity of Berkeley Puppet Interview scores are adequate. Criterion validity is adequate especially for the externalizing scales scores and to a lesser degree for the internalizing scales scores. However, due to its bi-modal distribution, examining reliability and construct validity is complicated, and therefore hard to investigate thoroughly.

• Psychologically controlling parenting as perceived by children was related to child reported internalizing and externalizing problems in young children. This association remained significant, even when controlling for important parenting styles as perceived by children, such as behavioural control and responsiveness.

• Parent reported maternal separation anxiety was related to child reported separation anxiety in young children cross-sectionally, but not longitudinally. Moreover, child perceived psychologically controlling tactics aimed at keeping children close (i.e., dependency-oriented psychological control) did not mediate the relation between maternal separation anxiety and separation anxiety in children, nor within or over time.

This project was primarily designed as a validation study of the SDQ. Therefore, we included a large cohort of participants which provides the advantage of increased power in statistical analyses. This enabled us to investigate to what extent contextual factors are related to problem behaviours in young children. As a means to this end, we employed the Strengths and Difficulties Questionnaire (SDQ) and the Berkeley Puppet Interview and investigated its psychometric properties. We investigated psychometrics of the Berkeley Puppet Interview in a subsample. The choice for a

Parent and teacher report versus a child self-report instrument was made based on the observation that investigating psychopathology in young children always implies using multiple informants to assess the problem behaviours of the child. An advantage of the inclusion of multiple informants is that it allows for a comprehensive assessment of problem behaviours. In this concluding chapter, first, we will reflect on our findings regarding psychometrics of the SDQ, next we focus on contextual factors in relation to problem behaviours in young children. Subsequently, we discuss these two issues (i.e., psychometrics and contextual factors) pertaining to the investigations we did using the Berkeley Puppet Interview. An advantage of these studies wherein child self-report was used is that it enables for the investigation of child perceptions that are not influenced by informant agreement issues. We end with a discussion on informant issues in research on child problem behaviours. Of note, all the reported findings are based on studies characterized by an observational design and therefore no causal inferences regarding any of the associations presented in this dissertation can be made.

Part I: Reflection on the main findings

Psychometrics: just a methodological nuisance?

Reliability: Alpha and Omega. In Chapters 3, 4, 5 and 7 we showed that the use of McDonald’s omega yields more precise and appropriate reliability indices than the traditionally used Cronbach’s alpha. These results do not merely show that we can ‘increase’ reliability indices, but reflect an alternative way of thinking about what reliability is exactly. To understand how social scientists used to conceptualize reliability, it is informative to return to the commonly used Cronbach’s alpha, an indicator of reliability that has dominated the field for decades (Schweizer, 2011). Alpha is an index of consistency, and is based on the premise that items representing the same construct should stimulate a consistent way of responding. Therefore, alpha represents the degree of consistency in response to items and is known to increase when participants respond consistently to items. Cronbach’s alpha was developed within classical test theory under the assumptions that a) items refer to a single factor (i.e., are unidimensional or homogeneous), b) errors are uncorrelated, and c) responses are normally distributed. In practice, some problems are encountered with the use of alpha, namely that the assumptions of unidimensionality, uncorrelated errors and normal distributions are often not met (Revelle & Zinbarg, 2009; Schmitt, 1996; Sijtsma, 2009). The assumption of normality is almost never met in social sciences, where distributions are skewed by definition, for example when investigating psychopathology. These practices lead to underestimations of reliability as expressed by alpha (Revelle & Zinbarg, 2009). The second problem is that many
researchers believe alpha is a measure for the unidimensionality of a questionnaire, evidenced by statements such as ‘as the alpha of .88 shows internal consistency is adequate’ (Sijtsma, 2009). As psychometricians have repeatedly shown, a high alpha can either be found in a questionnaire that is unidimensional (i.e., items refer to a single factor) and in one that is multidimensional (i.e., items refer to multiple factors) (Sijtsma, 2009). Therefore, one has to conclude that alpha just does not say anything about dimensionality of items. Put differently, alpha has little or nothing to do with the factor structure of an instrument.

Given these problems, alternative types of reliability have been developed, including McDonald’s omega (Revelle & Zinbarg, 2009). Importantly, these alternative types of reliability are developed under congeneric test theory where unidimensionality of items is a central tenet (Lucek, 2005; McDonald, 1999; Raykov, 1997). As such, an alternative way of thinking about reliability is introduced where, instead of consistency, unidimensionality is considered most important (Schweizer, 2011). Therefore, in order to examine reliability, the first step is to analyze whether the items are really unidimensional (i.e., whether the items refer to one latent construct). A means to testing whether items are unidimensional is factor analysis, which is usually done within a structural equation modeling (SEM) framework. Indeed, congeneric test theory stresses model fit of the congeneric measurement model. Next, after unidimensionality has been tested, reliability indices are based on factor loadings on the relevant dimension. The advantage of employing congeneric test theory, and thus working in a SEM framework, is that the assumption of unidimensionality is not violated and that corrections to the violation of the assumption of normality can be made. Moreover, congeneric test theory integrates the examination of validity and reliability, as reliability is based on factor analytical results. As such, reliability is now increasingly defined as an index for unidimensionality instead of consistency of items. This is not to say that the use of alpha is wrong. This is the case when a) alpha is applied to instruments which scores yield a normal distribution and b) when researchers do not interpret alpha as a measure for the underlying factor structure. This being said, there are indexes which yield more precise and appropriate results, and which are in line with modern theory, such as congeneric test theory (Schweizer, Altmeyer, Reiß, & Schreiner, 2010), and analysis techniques such as structural equation modeling.

How does this state of affairs compare to the literature on the SDQ? We found three recent studies on the SDQ reporting on coefficient omega (Gómez-Beneyto et al., 2013; Ezpeleta, Granero, de la Osa, Penelo, & Domínech, 2013; Kóbor, Takács, & Urbán, 2013), although the majority of studies on the SDQ continues to use Cronbach’s alpha as a reliability index and not coefficient omega (e.g., Liu et al., 2013; Mieloo et al., 2013; Moriwaki & Kamio, 2014; Theunissen, Vogels, Wolff, & Reijneveld, 2013). As such, it can be argued that many researchers continue to guide their research by classical test theory instead of congeneric test theory, at least when it comes to reliability. This is interesting, given that various studies seem to employ congeneric test theory when investigating construct validity, as evidenced by use of confirmatory factor analysis (Becker, Woerner, Hasselhorn, Banaschewski, & Rothenberger, 2004; Moriwaki & Kamio, 2014; Niclasen et al., 2008; Sijtsma, 2009). Therefore, one has to conclude that alpha just does not say anything about dimensionality of items. Put differently, alpha has little or nothing to do with the factor structure of an instrument.

As stated above, many researchers already analyze their main results according to congeneric test theory. Still, they seem reluctant to apply congeneric test theory when investigating the instrument they use to base their results from their main analyses on. It is unclear what may explain the tendency to use both theoretical perspectives, and to use these inconsistently. Although speculative, this practice may reflect that the fields of psychometrics and more content-driven subjects (e.g., the study into predictors of internalizing problems) are not highly connected. Evidently, these fields could benefit greatly from each other. The question is whether psychometricians should be more clear as to what the advantages are for social scientists regarding new developments in psychometrics, or whether social scientists should be more clear in what they need from psychometricians in order to analyze their results in a reliable and valid way, and in accordance with the most recent psychometric insights. Probably both apply.

**Construct Validity and Measurement Invariance.** In Chapters 3 and 4 we confirmed the five theoretically hypothesized scales of the SDQ, for both the parent and teacher version. Also, we replicated our findings by repeating the factor analyses across three measurements, which yielded similar results. These findings attest to the robustness of the factor structure, and confirm recent studies that also employed confirmatory factor analysis (Becker et al., 2004; Moriwaki & Kamio, 2014; Niclasen et
we show that the parent version of the SDQ is invariant across gender, age, ethnicity, maternal age and educational level, and survey method. Further, in Chapter 4 we showed that the teacher version is invariant for the strictest type of measurement invariance for age and ethnicity, but not for gender. Apart from the studies discussed in Chapter 3, we found only one recent study investigating measurement invariance for the parent and teacher version of the SDQ (Niclasen et al., 2013). These authors found the SDQ to be invariant across gender. For the teacher version specifically, we found two studies that tested measurement invariance (Ruchkin, Koposov, Vermeiren, & Schwab-Stone, 2012; Zwirs et al., 2011), respectively for adolescents and primary school children. Importantly, results from these studies are not consistent. Measurement invariance was established across gender and ethnicity by Zwirs et al., while our findings point to lack of measurement invariance for gender.

All in all, the paucity of studies on measurement invariance poses a threat to the evidence-base on construct validity of the SDQ. Findings from this thesis add to this evidence-base by showing that, for the Dutch parent and teacher version and for young children, meaningful group comparisons can be made. However, these findings are not exhaustive and call for more attention to measurement invariance in future research. Given that the SDQ was developed in the United Kingdom, it may be particularly likely that the SDQ was influenced by western views on what psychopathology constitutes. SDQ items thus might have a very different meaning to members of non-western societies and groups (Chen, 2008). As The Netherlands is a multi-cultural society, it is imperative that measurement invariance is established regarding various cultural groups. Although we showed that the SDQ is invariant across ethnicity, we did not distinguish between different cultural groups, such as children from Moroccan- or Turkish descent. This may be particularly important to investigate, as the meaning of the SDQ may be expected not to be similar for all groups of different ethnicity than the Dutch one. Instead, it may be expected that the SDQ may mean something else for each cultural group, and this should be tested. Furthermore, this thesis was directed at younger children, but establishing measurement invariance for older children and adolescents is direly needed as the SDQ is used extensively in research on children and adolescents. Testing measurement invariance is consistent with congeneric test theory, in that unidimensionality of items is extensively investigated by testing measurement invariance for certain constructs or groups.

Concluding, congeneric test theory is a more holistic and modern approach to testing psychometric properties of an instrument, than classical test theory. The two most important advantages of employing congeneric test theory are that a) the assumption of unidimensionality is not violated, and b) that this approach provides solutions for the violations of normality as these are available within a SEM framework. Also, this theoretical perspective guides researchers as to what psychometrics are the most important psychometric aspects to investigate (i.e., reliability and construct validity), because of its focus on unidimensionality instead of consistency (Evers, Lucassen, Meijer, & Sijtsma, 2010). As such, researchers are encouraged to simultaneously test reliability and construct validity. Finally, this perspective enables for integration of reliability and construct validity, by examining these properties simultaneously. This is opposed to current practices where it seems to be the case that when investigating reliability, researchers still use classical test theory to guide their choices, while violating its assumptions frequently. As a consequence, the ubiquitously reported Cronbach’s alpha is usually an underestimation of reliability.

**Criterion Validity: Concurrent and Predictive Validity.** In Chapter 4, we replicated findings from previous studies that the SDQ’s criterion validity is adequate when examined concurrently. Also, we showed that both the parent and teacher version of the SDQ are predictively valid when it comes to some outcomes. Predictive validity for the parent version was demonstrated by showing that developmental pathways of the SDQ are predictive of scores on inadequate parenting and daily hassles. Predictive validity for the teacher version was demonstrated by showing that developmental pathways of the SDQ are predictive of the extent to which children are liked by their classmates (i.e., likeability). All in all, we may conclude that SDQ scores reported for young children by parents are predictively valid, concerning inadequate parenting and daily hassles and for teachers concerning likeability. The predictive validity of the SDQ has been largely overlooked in the very large literature on this instrument, and specifically concerning the parent and teacher version and younger children. Two studies were found that reported on stability of SDQ scores for young children (Perren, Stadelmann, Von Wyl, & Von Kitzing, 2007; Hawes & Dadds, 2004), which is framed as predictive validity in these papers. These studies indeed found that SDQ scores predict SDQ scores one year later, for both the parent and teacher version. Though, as the definition of predictive validity is ‘the ability of an instrument to predict scores on another criterion measure in the future’, these studies did not really test whether the SDQ is predictively valid. Clearly, findings from this thesis add...
to these findings as we followed children during three years, and compared SDQ scores to parenting and peer factors.

Conclusion regarding psychometrics
As for psychometric properties of the SDQ, findings from Chapters 2-4 indicate these have been found adequate, in two different large samples in young children, and when screening for psychopathology in a community sample. These three clauses concerning the large sample, age of the children and screening in a community sample are particularly important. These issues are important to stress because they prevent making statements based on this thesis, such as ‘the SDQ is a good instrument’ or ‘because the SDQ is a good instrument, high SDQ scores mean that a child has a mental illness’. Although these statements may seem strange at first glance, it is no exception that scores from instruments are interpreted as such, as if these scores entail some sort of truth. Evidently, these statements are generalizations that cannot be made based on our findings. Indeed, it is important to observe that an instrument does not possess psychometric properties, but that these psychometric properties are dependent on the purpose and the population (Hunsley & Mash, 2007). This is not to say that the SDQ should not be employed, to the contrary, psychometric research should facilitate the use of the SDQ. Still, we do caution against the interpretation of SDQ scores in individuals with minimal education in statistics as this poses a risk to a comprehensive interpretation of SDQ scores. Therefore, we encourage the use of the SDQ based on our findings for screening purposes, thus using low-risk community samples. Also, professionals with a sound education in statistics should interpret scores of the SDQ and should make nuanced statements concerning these scores. These considerations for use are elaborated below. Finally, it is up to researchers to investigate an instrument meticulously regarding its psychometric properties before using an instrument in content-driven studies. Thus, as we argued above, psychometrics are not just a methodological nuisance, and should not be, especially so to researchers and professionals who use the SDQ.

Contextual factors in relation to internalizing and externalizing problems
As described in the introduction of this dissertation, problem behaviours reside in a social context. Studying problem behaviour in relation to these contextual factors thus seems a fruitful avenue. In Chapters 5, 6 and 7 we examined interrelations between internalizing and externalizing problems and the extent to which contextual factors were related to these problems in young children. In Chapter 5 we found that internalizing and externalizing problems did not affect each other over time (i.e., heterotypic continuity). We did find that previous levels of internalizing problems were strongly associated with later internalizing problems, and similarly so for externalizing problems, indicating strong homotypic continuity. Also, we examined predictors of child psychopathology and sought to answer whether these predictors would be responsible for the associations over time between internalizing and externalizing problems. As we did not find any cross-lagged associations, this question became redundant. We did find that maternal health was associated with internalizing problems and parenting stress with externalizing problems over time and while controlling for concurrent relations.

In Chapter 6 we found that internalizing problems cluster in reciprocated but not in unilateral friendships. Importantly, we showed that this association could not be explained by concurrent externalizing problems. These findings extend previous research on homophily of internalizing problems in friendships, as this association was primarily demonstrated in adolescence (Brendgen, LaMarche, Wanner, & Vitaro, 2010; Hogue & Steinberg, 1995; Prinstein, 2007; Rubin, Woslawowicz, Rose-Krasnor, Booth-LaForce, & Burgess, 2006; Stevens & Prinstein, 2005), with one exception (Mercer & DeRosier, 2010). Also, these findings underline the possibility of friendship with a similar peer as a risk factor for internalizing problems. Indeed, it has been shown that socially withdrawn adolescents reported their friends to be less helpful, less intimate and provided less guidance (Rubin et al., 2006), although this association could not be found in children aged 10-12 (Klima & Repetti, 2008), and has not been tested in younger children. One study did show that friends aged 8-10 socialized each other into higher levels of internalizing problems over time (Mercer & DeRosier, 2010). Still, the mechanisms by which this process occurs remain unclear. Also, the question whether children with internalizing problems befriended children with internalizing problems intentionally, or whether they unintentionally end up with those children (i.e., selection processes) remains open.

Another question that remains open and is not addressed regularly in the literature is how friendships of children with internalizing problems relate to broader aspects of the peer group, such as victimization, peer rejection (often expressed as low social preference), and popularity. This is quite surprising, as it has been shown that peer rejection and victimization are related robustly to internalizing problems both within and over time (see for a meta-analysis, Reijntjes, Kamphuis, Prinzie, & Telch, 2010; Van Lier & Koot, 2010) and also to difficulty in forming friendship and low quality friendships (Brendgen, Little, & Krappmann, 2000; Pedersen, Vitaro, Barker, & Borge, 2007; Rubin et al., 2006). Recently, the selection and socialization processes through which children form friendships have been linked to popularity in the peer group, and partially explained how popularity emerged (Dijkstra, Cillessen, & Borch, 2013). Perhaps, this integration of the study of friendship within the broader peer group could be a fruitful approach in the study of internalizing problems as well. Our results even hint to a further broadening of the peer group to microsystem factors, such as parenting, being still more fruitful. We found low social preference being
related to internalizing problems both within and over time on a bivariate level, but not when we controlled for inadequate parenting, maternal health, parenting stress and externalizing problems (Chapter 5). These results are particularly important as in most studies examining peer relations and internalizing problems, associations are studied in relative isolation of contextual factors, such as parenting, although these studies did take other contextual factors into account, such as concurrent problem behaviours (Spittal, Van Der Leij, Onghena, & Colpin, 2013; Reijntjes et al., 2010; Van Der Leij & Koot, 2010). Evidently, this impedes the results of these studies, as results regarding peer relations may be, in part, attributable to contextual factors.

In Chapter 7 we investigated the developmental course of internalizing and externalizing problems, and parenting stress. Regarding internalizing problems, we did not find a decrease or increase of these problems during early childhood. Regarding externalizing problems, we found these problems to decrease during early childhood. These results mesh with previous studies consistently reporting that externalizing problems decrease during childhood (e.g., De Haan, Prinzie, & Dekovic, 2010; Maughan, Collishaw, Meltzer, & Goodman, 2008; Prinzie, Onghena, & Hellinckx, 2006). As for parenting stress, we found that this decreased during early childhood, which is in accordance with a previous study (Williford, Caflkins, & Keane, 2007).

Second, we investigated whether parenting stress and internalizing and externalizing problems were interrelated. Decreases in parenting stress were related to decreases in externalizing problems, and decreases in parenting stress were related to internalizing problems over time. As for directionality, we found support for externalizing problems impacting parenting stress, and vice versa. Regarding internalizing problems, we only found support for parenting stress impacting internalizing problems. Thus, while for externalizing problems both a child effects and parent effects model apply, for internalizing problems, we only found support for a parent effects model. When looking at findings from Chapter 5, we found parenting stress to be related concurrently to internalizing problems, but not over time. Regarding externalizing problems, we found parenting stress to be related to externalizing problems both concurrently and longitudinally (Chapter 5). Taken together, these findings provide evidence for transactional relations between externalizing problems and parenting stress (Cicchetti, 2006). In contrast, regarding internalizing problems, we found no evidence for transactional relations. Hence, results indicate that internalizing and externalizing problems are characterized by a differing developmental course, and with specific factors being related to each of these problem clusters.

Microsystem factors related to internalizing problems? In this thesis, we investigated both internalizing and externalizing problems. Two problems were encountered in the study of internalizing problems in childhood. First, relative to the study of internalizing problems, much more research has been conducted into studying externalizing problems (e.g., Deater-Deckard, 2001). Second, relative to the period of childhood, much more research has been conducted into studying internalizing problems during adolescence (Zahn-Waxler, Klimes-Dougan, & Slattery, 2000). Given this shortage of attention for childhood internalizing problems, we focus this general discussion on our findings concerning internalizing problems. The reason why fewer studies have investigated what risk factors are associated with internalizing problems during childhood than studies conducted during adolescence is probably twofold. First, internalizing problems seem to cause less direct problems in comparison to externalizing problems to parents, teachers and peers. Second, more variability in internalizing problems is found in adolescence due to its increase during this period. Also, a pragmatic reason seems to be that studying the developmental period of adolescence seems to be more practical because it allows for the use of self-report and a direct assessment of adolescent’s perception of problem behaviours and related factors. Still, researchers may miss important clues as to the development of internalizing problems by focusing solely on adolescence. Evidently, internalizing problems during adolescence do not appear out of the blue and therefore, it was deemed important to investigate what factors related to internalizing problems can be identified during childhood. For example, internalizing behaviours during childhood may be part of a chain of events that eventually escalate into serious depression or anxiety disorders later in life (e.g., adolescence). Taken together, findings from this thesis show that there are several concurrent risk factors for internalizing problems in early childhood that all reside within the family or the peer group (i.e., microsystem factors). These factors are externalizing problems and having a friend who also has internalizing problems. Risk factors that are related over time identified in this thesis are previous levels of internalizing problems and maternal health.

However, how these factors interact to increase internalizing problems and what processes underlie the development of internalizing problems is not well understood, due to a lack of studies investigating how these factors interact (Mills et al., 2012; Degnan, Almas, & Fox, 2010). Although we looked at the development of internalizing problems in a developmental period prior to adolescence, by using both cross-sectional and longitudinal designs and taking into account risk factors comprehensively, we did not find stable patterns of how risk factors interacted to impact internalizing problems. It has been argued that indeed aversive parenting may not be strongly related to internalizing problems in the absence of cumulative effects (Paulussen-Hoogeboom, Stam, Hermanns, & Peetsma, 2007), such as interactions between parental overcontrol and inhibited temperament. Also, reviews show that that only a small part of internalizing problems may be explained by parenting (McLeod, Weisz, & Wood, 2007; McLeod, Wood, & Weisz, 2007). However, it has also been shown that the genetic component in childhood anxieties are not very high
(Eley, Rijndijk, Perrin, O’Connor, & Bolton, 2008). Therefore, it remains unclear what factors are associated with internalizing problems. These findings call for a more integrated approach for studying internalizing problems (cf. Degnan et al., 2010). However, individual researchers often focus on a specific domain, and become specialized in this domain, which leads to fragmented research (Hollenstein & Lougheed, 2013). This probably has impeded integration of findings from temperamental and environmental research on internalizing problems in childhood.

**Limitations and directions for future research**

The results in the current dissertation provide clear new insights into reliability and validity issues. In addition, the results regarding internalizing problems provide a starting point for future research. As it is deemed important that studies on internalizing problems in childhood are continued in future research, this section is primarily directed at avenues for future research regarding the study of internalizing problems. Based on the results of our studies on internalizing problems we recommend two types of future studies. First, large-scale longitudinal studies are needed to investigate the long-term course of internalizing problems during childhood and contextual factors related to these problems over time. Second, small-scale studies are needed in order to elucidate underlying mechanisms by which internalizing problems develop in childhood. While describing these future directions, we also discuss the limitations to the studies presented in this dissertation.

With regard to large-scale longitudinal studies, these studies should be aimed at identifying the developmental course of internalizing problems. As such, it can be examined what the typical course of childhood problems is (Sroufe, 1997), such that it can be determined when children deviate from what is typical (Cicchetti & Rogosch, 2002). Although important work has been done on this topic (Sterba, Prinstein, & Cox, 2007; Côté et al., 2009), it remains unclear whether internalizing problems increase (Côté et al., 2009; Bongers, Koot, Van der Ende, & Verhulst, 2003), decrease or remain stable (Maughan et al., 2008; Kelley, Loffhouse, Bates, Dodge, & Pettit, 2003; Sterba et al., 2007) during childhood. Two of these studies were conducted very early in development, starting from age 1.5 and following children until age 6 (Côté et al., 2009; Sterba et al., 2007). If anything, these studies show that internalizing problems are already present during the preschool years. Moreover, development in the early years, age 0-4, is characterized by rapid and large changes, which are reflected in change in structure, increased function, and explosive growth of brain volume (Casey, Tottenham, Liston, & Durston, 2005). Therefore, in order to disentangle what is normative and what is deviant from the norm, we argue that large-scale longitudinal studies should be conducted as early in development as possible. In order to capture the large and rapid changes observed in the preschool years it is imperative that study intervals are matched to these changes. Ideally, the younger the child, the shorter the intervals between measurements should be. Currently, intervals between measurements of 6 or 12 months are typical in longitudinal studies. When studying newborns until preschool it would be more informative to reduce those intervals to monthly or weekly intervals. Of course, such studies would be very time-consuming and expensive to conduct. Still, one could argue that studies wherein development is not fully captured can be considered even more expensive, as these may lead to incomplete results. Concluding, to study the course of internalizing problems we argue that large-scale studies are needed, beginning as early as possible and measuring behaviours as often as possible.

Evidently, psychometrically sound questionnaires should be administered to multiple informants in such a large-scale study. In keeping with recommendations derived from informant discrepancies research (De Los Reyes, 2013), it is essential in such a study that researchers measure the same construct with multiple measures to ensure the validity of this construct and that the conclusions of the study should be drawn specifically per informant. Also, in order to ensure variability in internalizing problems, groups with a low socioeconomic status should be included in large-scale studies. Finally, from a methodological viewpoint, different research questions may be addressed in such large-scale studies. First, following a variable-centered approach, the most important research questions are: What is the (mean) onset of internalizing problems in children? How does this onset relate to (mean) internalizing problems over time? In such a variable-centered approach it is assumed that all children follow a similar developmental trend (Von Eye & Bergman, 2003). From a developmental psychopathology perspective, however, it would be more informative to study qualitative form of the development of internalizing problems (Cicchetti & Rogosch, 2002). Moreover, it is assumed that children’s problems are heterogeneous. From such a person-oriented approach, the most important questions are: Which children continue to develop internalizing problems, and which children discontinue? Can subgroups of children with differing levels of internalizing problems be identified, and what do those subgroups look like? Can these subgroups of children be identified over multiple time points, and is membership in these subgroups stable or instable? A limitation to the studies presented in this dissertation regarding the development of internalizing problems are selection bias and the relatively high rates of attrition in our sample. Regarding selection bias, we did not randomly select children from the population to participate. As for maternal educational level, mothers in our sample were somewhat higher educated than mothers from the Dutch population. Still, children from our sample were by and large comparable to children from the Dutch population regarding ethnicity, family composition, and marital status. Ideally, researchers should select their sample by stratifying and at-random selection to prevent selection bias taking place. Regarding attrition, this may have led to only more healthy and well-functioning (i.e., high SES) participants participating in all three
exactly, and what happens in the child’s proximal environment when a child shifts
and explaining intra-individual change instead of inter-individual change, and utilizing
person-centered instead of variable-centered designs. Thus, change relative to the
individual is studied instead of change relative to the other participants in the sample.
A limitation to the studies presented in this thesis is that our measurement of internalizing problems was not
elaborate, with the SDQ’s emotional symptoms scale containing only five items. Further, we focused on questionnaires, with has the disadvantage of focusing too
much on perception, while not taking more objective measures into account. Clearly,
taking a multi-disciplinary approach, where several methodologies such as questionnaires, observations, interviews and peer ratings are used is preferable as the
combination of these methodologies is thought to yield the most coherent picture
of internalizing problems (Degnan et al., 2010). For example, from cognitive learning
theory perspectives (e.g., Field, 2006) it has been shown that information processing
styles including attention and interpretation biases are associated with internalizing problems in childhood (In-Albon, Kossowsky, & Schneider, 2010; Puliafico & Kendall,
2006). Also, from a biological perspective, it has been shown that genetic factors,
(neuro)biological and physiological regulatory processes are associated with internalizing problems in childhood (Gregory & Eley, 2011; Guerry & Hastings, 2011;
Meyer et al., 2013). Finally, from a contextual perspective, studies have shown that
parental control is related to internalizing problems in childhood (Chapter 9; Van der
Bruggen, Stams, & Bögels, 2008). Combining these perspectives may answer questions that have risen from this thesis regarding how internalizing problems are
affected by several factors.
Theoretically, such an approach is advocated in dynamic systems (Granic, 2005;
Hollenstein, 2011). In short, this perspective posits that development is constant and
should be viewed from many levels and time scales. As such, the interest is describing
and explaining intra-individual change instead of inter-individual change, and utilizing
person-centered instead of variable-centered designs. Thus, change relative to the
individual is studied instead of change relative to the other participants in the sample.
Also, within dynamic systems the degree to which individuals recover from certain
negative emotions to more positive emotions is considered very important (Granic,
2005; Hollenstein, 2011). For example, one could be interested in how fast, when
exactly, and what happens in the child’s proximal environment when a child shifts
from internalizing behaviours (e.g., crying, withdraw from interaction, worrying,
negative affect) to normative behaviours (e.g., playing, staying in the interaction,
normal or positive affect). As such, the following research questions could be
addressed. In whose company do children show internalizing problems most? What
is the nature of this company in terms of age, relationship (i.e., family, friend, peer),
personality and interaction style? When children do show internalizing behaviours,
how fast do they recover from these behaviours? What do children do when they
recover from internalizing behaviours, how do they recover? What behaviour of their
company is associated with more internalizing behaviours and what behaviour is
associated with less internalizing behaviours?

Implications
As the findings regarding psychometrics of the SDQ are the most clear when
considering Part I of this thesis, we focus this section on implications regarding the
use of the SDQ. Based on the aforementioned issues and findings, we recommend
taking the following into account when anticipating on using the SDQ. We differentiate
between practical, scientific, policy, and ethical considerations. We start with practical
considerations. First, SDQ scores are a global indicator of child psychopathology
and prosocial behaviour. Therefore, the SDQ is particularly suited for large cohort
studies or as part of screening for vulnerable children in terms of problem behaviour
in school settings. When used in a school setting, the SDQ should be administered
regarding each child in the classroom. This is advised because when the aim of the
SDQ is to screen, each child should be included in such screening in order not to
overlook any children. This advice is backed by the literature on confirmation bias
where people tend to confirm what they expect based on their values (Garb, 1998;
Kidc, 2006). Interpretation of SDQ scores should be performed by a professional with
a scientific education, specifically this professional should be aware of the following
a) that SDQ scores are a reflection of the perception of the informant b) that although
scores may seem absolute, they reflect the construct being measured at a certain
point in time c) this construct is derived from a theoretical perspective that has
influenced what is regarded important to measure, and d) that scores derived from
questionnaires should always complement a broader assessment, such as
observation of the child or an interview with the child or its parents.

As for scientific implications, when the SDQ is used in research its factor structure
should always be investigated using CFA. Preferable, the factor structure should be
tested in each sample wherein the factor structure was not formerly tested. With
regard to reliability, coefficient omega should be used when reporting on reliability if
distributions deviate from the normal distribution. As described above, this must be
the case when using the SDQ, as the SDQ is a measure for psychopathology and
therefore its scores are skewed by definition. Therefore, it is warranted that coefficient
omega is used when reporting on reliability of the SDQ. When examining group differences is the aim of the study, researchers should be aware of measurement invariance. When measurement invariance is not established for a certain variable (e.g., culture) researchers should refrain from comparing groups regarding this variable. Ideally, researchers should first test for measurement invariance, if this is not yet established, and subsequently answer their main research question. On the other hand, we acknowledge that this methodological endeavour is time-consuming. This makes testing of measurement invariance less attractive. When researchers have good reasons not to test for measurement invariance this could be articulated in their study, such that it is clear to the reader why the authors did deem it important to test for measurement invariance. Also, in longitudinal studies, the meaning of a construct may change over time, warranting investigation of measurement invariance. Finally, investigating measurement invariance is important in prevention and intervention studies, as the construct may change due to treatment effects (Chen, 2008).

With regard to policy, in the Netherlands test users may look up information on psychometrics of tests via several organizations, such as via the Dutch Youth Institute (NJI; www.nji.nl). The authority among these organizations is the Committee on Test Issues Netherlands (COTAN; www.cotandocumentatie.nl; Evers et al., 2010), whose goal is to judge the quality of tests in terms of psychometrics and to inform users about the quality of tests. The quality of the test is published on COTAN’s website, which is accessible to subscribers. COTAN publishes a summary of its judgment on its website in terms of psychometric properties being inadequate, adequate or good, as well as a more detailed report. With this practice it seems as if psychometric properties are static and not dependent on the sample and purpose of the test. We argue that it should be made more clear to the test user under which conditions the psychometric properties have been investigated, and thus under which conditions it is justified to use the test (cf. Hunsley & Mash, 2007). A short description of these conditions and an advice to the test user would be particularly helpful in this sense. This being said, in comparison to other European countries, the Dutch system of COTAN is considered of very good quality, given its regular updates of psychometric properties considered important and use of modern methodological theories (Evers, Sijtsma, Lukassen, & Meijer, 2010).

These considerations being made, the question remains whether the SDQ should be used at all as a screening instrument. Surprisingly, this question is not posed much in the social sciences, although it seems important to ask ourselves whether there really are only advantages to screen children for problem behaviours. To our best knowledge, very little if nothing, has been written on this ethical issue in the field of the social sciences. The available literature on screening comes from the medical sciences, wherein such dilemmas are considered, for example in screening for breast cancer (Gates, 2001). So, should the SDQ be employed as a screening measure, for instance to screen all children at elementary school for problem behaviours? A reason why this is advocated is because these children could be helped earlier in their development, when their problems are not yet integrated fully into their personality (Costello, Egger, & Angold, 2005). Also, it is regarded important that screening is quick, and therefore the screening instruments should be as short as possible. Although this reasoning seems morally adequate, it could be that we are doing more harm than good by screening these children. First, because screening is usually initiated by a professional, there is a so-called ‘implied promise’ that screening will do more good than harm (Malm, 1999; Marshall, 1996). However, this is not always the case. A potential harm lies in the false-positive result, where a child is identified as having problems, while this is actually not the case. Such an identification as a child being a ‘problem-child’ may lead to anxiety and worry not only in the child, but also in his/her parents (Marshall, 1996). Second, false-negative results may be harmful, where a child is considered not a ‘problem-child’ while there are actually problems the child faces. This may encourage children and parents to ignore those problems. Finally, a screening instrument may correctly identify a child as having problems, but if the therapy is ineffective or harmful, the child has been harmed rather than helped (Marshall, 1996). These disadvantages of screening may be explained further by giving an example. What if a teacher would fill out a questionnaire for all children in her class as part of a general mental health screening procedure. She would find out that a certain child scored above the clinical cut-off. Suppose the teacher would interpret this report as the child having a serious problem, and would confront the parents with it. What if the parents would not recognize these problems? What if they do not experience problems with their child and what if also the child itself would not? The question is what to do next. Find help for the child, or leave things as they are? Whatever choice is made in this phase, the measurement has already done its work in changing the perception of the teacher, and possibly also that of parents. The question is whether the teacher and the parents are able to not see this child as a problem child anymore. Because it is not unlikely that this example will actually take place in practice, we should take the greatest caution in interpreting scores generated by a questionnaire. Only when these strict guidelines on using the SDQ can be followed, we advise the use of the SDQ as a screening instrument. As such, it is evident that a focus on speed instead of a focus on quality seems not the best option when evaluating a child’s health or problem behaviours (Dehue, 2008; Verhaeghe, 2012).
Part II: Reflections on the main findings

The project the results of part II were based on, came from a sub-sample of our large validation study as described above. In the second part we studied the associations between parental control and childhood problem behaviours. A strength of this second part is that we were able to study psychologically controlling parenting in young children by using an innovative approach to child self-report. Besides, we also examined how well this child self-report instrument performed in terms of psychometrics. Again, all presented associations are correlational and therefore no causal inferences can be made based on our findings.

Child self-report
In Chapter 8 we showed that BPI scores on its psychopathology scales relate to each other over time, attesting to adequate stability. Further, BPI scores intercorrelate and correlate to scores on similar instruments, as reported on by parents and teachers, which serves as evidence for its congruent and concurrent validity. We are among the first to show this for the Dutch version of this instrument (Ringoot et al., 2013). These findings mesh with previous studies on psychometrics of the BPI (Ablow et al., 1999; Arsenault, Kim-Cohen, Taylor, Caspi, & Moffitt, 2005; Luby, Belden, Sullivan, & Spitznagel, 2007). However, important aspects of this interviewing method, its internal reliability and construct validity, have not been extensively investigated. Moreover, the available studies show mixed findings concerning reliability and construct validity (Ablow et al., 1999; Ringoot et al., 2013). This may have to do with the bimodal and non-normal distribution of BPI scores, which does not meet the assumptions of the most common statistical tests (i.e., assume normality of data). Although beyond the scope of this thesis, it is deemed important to find a solution for this issue.

What is clear from decades of research, is that child mental health assessment requires multiple informants (Hunsley & Mash, 2007). Whereas children were formerly viewed as not being able to report on their own feelings of distress, it is now widely agreed that children can be reliable informants and as such contribute to the assessment of problems (Arsenault et al., 2005). Nevertheless, child self-report is still regarded complex due to children’s proneness to social desirability (Grills & Olleandick, 2002). With the inclusion of young children as informants, the question of how to handle child self-reports when these diverge from what teachers and parents report regarding these children also comes to mind. This complicates the child self-report issue even further, as integrating data from multiple informants is complex. So, although it is acknowledged by the scientific public that multiple informants are required in child research, it seems to be hard to analyze results from a multi-informant perspective (De Los Reyes, 2011). It is considered complex to integrate data from multiple informants because these reports more often diverge than they converge, yielding complex decisions to be made as to how to analyze those diverging reports. How to handle reports from multiple informants is elaborated in the implications section.

Psychologically controlling parenting in relation to internalizing and externalizing problems
In Chapters 9 and 10 we examined whether psychologically controlling parenting is related to child problem behaviour, and whether dependency-oriented psychological control explained relations between maternal and child separation anxiety. We showed that child perceived psychological control was robustly related to child reported internalizing and externalizing problems, as we showed this while controlling for child perceived responsiveness and behavioural control (Chapter 9). While this association has been repeatedly demonstrated in adolescents (e.g., Barber, 1996; Barber, Stolz, & Olsen, 2005; Gray & Steinberg, 1999; Pettit, Laird, Dodge, Bates, & Criss, 2001; Soenens, Vansteenkist, Luyten, Duriez, & Goossens, 2005; Soenens et al., 2008), we were the first to show psychological control is directly related to maladaptive outcomes in children. Also, we examined dynamics of psychological control, as we examined whether a characteristic of controlled intrapersonal functioning, separation-anxiety in mothers, would be related to separation anxiety in children via psychologically controlling parenting. Although we found a marginal relation between parent reported maternal separation-anxiety and child perceived dependency-oriented psychological control, and a concurrent association between child perceived dependency-oriented psychological control and self-reported separation anxiety in children, we did not find mediation (Chapter 10). Also, we did not find that parent reported maternal separation anxiety was associated with self-reported separation in children over time.

Together, these findings point to child perceived psychological control being important within, but not over time during early childhood. As such, these findings mesh with research showing concurrent associations of adolescent perceived psychological control and maladaptive outcomes in adolescence (Soenens et al., 2005). Though, the question remains why perceived psychological controlling parenting is associated with internalizing problems over time in adolescence, even when controlling for previous levels of internalizing problems (Soenens et al., 2008), but not in childhood. As outlined by Self-Determination Theory (SDT; Deci & Ryan, 2000), the mechanism by which psychological control exerts its influence on problem behaviours, is through stifling the need for autonomy. That is, psychological control would hinder children in developing a healthy sense of self, wherein they feel autonomously motivated to engage in certain behaviours. It is argued that if psychological control thwarts the need for autonomy, it should impact development regardless age. However, this is not to say that the way autonomy frustration manifests is age-invariant. While adolescence is characterized by increased independence and
CHAPTER 11

Limitations and future directions

The results in this dissertation provide new insights into reliability and validity issues of the Berkeley Puppet Interview. In addition, the results regarding psychological control are innovative in showing that psychological control already manifests as a risk factor for internalizing and externalizing problems in developmental periods prior to adolescence. We recommend three primary directions for future research. While describing these future directions, we also discuss the limitations to the studies presented in this dissertation.

First, the one-year interval between the BPI measurements seems somewhat too long, as parenting and child development are considered to be processes that continually change and impact each other (Granic, 2005; Hollenstein, 2011). The one-year interval in our study was mainly chosen for pragmatic reasons, but based on the statement made above it would be more sensible to study developing children at more time points than two, and in closer time frames. A particularly suited methodology for studying problem behaviours in conjunction with parenting seems to be ecological momentary assessment (EMA; Shiffman, Stone, & Hufford, 2008). In EMA participants are repeatedly assessed concerning thoughts, feelings and behaviour in their natural environment. Still, it would be impossible for children aged 4-5 to report on those thoughts, feelings and behaviour themselves as this requires reading skills that children this age do not yet possess. Possibly, children aged 6-7 could provide this information when assisted with age-appropriate language and signs, which may be facilitated quite easily by digital media (Granic, Lobel, & Engels, 2014). As such, it may become clearer when children report on low mood and what factors are associated with such low mood. Subsequently, by studying both children and parents in such a framework, it may become clear how children and parents react to each other, and how they interact when children experience internalized problems.

Second, the reliability and the factor structure of the BPI could not be thoroughly investigated due to its bi-modal distribution. Although some researchers apply factor analysis when investigating the BPI (Ablow et al., 1999; Ringoot et al., 2013), we argue that this is not the most promising avenue of research, as the BPI’s scores do not fit with the assumptions of these statistical tests. Therefore, we argue for re-development of the BPI wherein either BPI scores fit with assumptions of statistical tests (i.e., scores are distributed normally) or BPI scores can be analyzed differently such that such an adjustment of the BPI is not necessary. As described above, child self-report remains complex, mainly due to issues of social desirability. Therefore, such a development should include a pilot version of this new instrument, wherein it is examined whether children feel obliged to answer in a certain direction. Evidently, the issue of the bi-modal distribution of BPI scores should also be resolved in this development. Not using two puppets would resolve this, although by doing this the strength of the BPI, namely that the child can identify with either puppet, is lost. Third, although the perception of psychologically controlling parenting seems to be important given its robust association with childhood psychopathology concurrently, its role is not yet entirely clear. Future research is needed where children are followed over time, but with shorter time intervals than a year, to study whether psychological control has a short-term impact on children’s problems. As such, it would also be important to investigate moderating factors, namely, whether associations between psychological control and internalizing problems are stronger for children who are already at-risk for internalizing problems. While it has been shown that the link between perceived psychological control and internalizing problems was moderated by difficult temperament in a cross-sectional study (Morris, Silk, Steinberg, Sessa, Avenevoli, & Essex, 2002), this has not yet been investigated in a longitudinal study. Therefore, the direction of these associations is not yet understood. Furthermore, while temperament is one of the factors rendering children vulnerable to psychological control, this is only the starting point of the search for moderating factors. For example, in adolescence it has been shown that self-worth and a relationally aggressive interpersonal style are
associated with internalizing problems in children of psychologically controlling parents (Soenens, Vansteenkiste, Goossens, Duriez, & Niemiec, 2008; Soenens et al., 2005).

Fourth, with longitudinal studies with short time intervals, it might be possible to identify a chain of events leading up to internalizing problems, i.e., to identify mediating mechanisms of psychological control and internalizing problems. Although need satisfaction has been suggested as the hypothesized mechanism underlying the link between psychological control and child problems (Soenens & Vansteenkiste, 2010), from this thesis it is unclear whether this is really the case. More research is needed in order to clarify whether need satisfaction is indeed the mechanism underlying the psychological control problem behaviour link. Age-appropriate measures of need satisfaction, such as the need for competence, autonomy and relatedness, should be developed in order to test this hypothesized mechanism. Further, it should be tested whether these measures indeed mediate between relations of psychologically controlling parenting and internalizing problems in childhood. Finally, as argued above, psychological control may not increase internalizing problems, but may increase factors related to these problems, such as contingent self-esteem, where individuals feel they have to fulfill criteria to feel good about themselves (Wouters et al., 2013). Investigating such proxies of internalizing problems thus might be a fruitful approach for studying the mechanism of psychologically controlling parenting and problem behaviours. This seems especially true for the study of childhood internalizing problems, where limited variability in these problems is often a problem in childhood-focused studies.

Finally, present it is not clear at what level psychologically controlling parenting is harmful. While associations of psychological control with problem behaviours have been reported consistently in adolescents at a correlational level, few studies have examined how strong the degree of psychologically controlling parenting should be to have a detrimental impact on both children and adolescents. As such, it remains unclear whether only subtle degrees of psychological control are already associated with problem behaviours, or that strong and pervasive degrees of psychological control are needed to associate with problem behaviours. Also, in such a view, it may be possible for researchers to study whether some children may be able to recover from psychologically controlling parenting. Concluding, it may be important to examine to what extent the degree of psychologically controlling parenting is associated with in- and externalizing problems in children. In addition to the knowledge this could bring about the nature of psychologically controlling parenting, such information would be particularly helpful for designing and targeting interventions for parents who employ psychologically controlling parenting tactics.

**Implications**

Based on our findings and the issues pertaining to these findings raised above, we describe implications of our findings. As our findings concerning psychologically controlling parenting are regarded preliminary, this section focuses on implications of using the BPI and informant issues in studying child psychopathology. First, for practice we recommend the use of the BPI as a qualitative instead of a quantitative measure, as we could not test the BPI’s factor structure adequately (Chapter 8). Therefore, we urge clinicians not to derive scores from the BPI, but to observe the interaction between the puppets and the child and to look for patterns in the child’s behaviour that coincides with the child’s problematic behaviour. As such a qualitative measure, the BPI is attuned to the child’s developmental level by engaging the child in a playful interaction (e.g., Eder, 1990). At the same time, using the BPI has an advantage over unstructured play activities with children because of its structured format which prevents clinicians from asking questions in one direction due to confirmation bias (Garb, 1998; Kida, 2006). Moreover, it helps children in not feeling pressured to answer what they think the clinician expects them to say (i.e., social desirability; Grills & Ollendick, 2002). When the BPI is re-developed such that its factor structure can be tested, we advise that the BPI can be used to extend evidence-based assessment. As such, the BPI may be used as a semi-structured interview which has the advantage of minimizing common biases observed during diagnostic processes where no structured interviews are used (Garb, 1998).

**Informant issues.** An issue relevant to all chapters in this thesis is a phenomenon termed ‘informant disagreement’. This phenomenon pertains to what has been called the most robust finding in clinical child research (De Los Reyes & Kazdin, 2005), namely, that informants tend to disagree with each other when reporting on psychopathology (De Los Reyes, 2011), and risk factors correlated to child psychopathology such as family conflict and parenting (e.g., Gonzales, Cauce, & Mason, 1996; Taber, 2010). Evidently, this issue complicates interpretation of results and obscures drawing clear conclusions. This may be illustrated by looking at simple correlations between reports of parents and teachers on the SDQ as reported in Chapter 4, and children and parents, and children and teachers, as reported in Chapter 8. These correlations hover around .20-.30, which is low to medium. So what may be concluded from these correlations? Is somebody not reporting ‘honestly’? Is somebody not ‘right’? Who is the best informant? Whose reports should we use in research? Are these correlations low due to measurement error? Despite the longstanding history of the informant disagreement phenomenon, as yet, there is no clear answer to some of these questions. It is becoming increasingly clear that the search for the optimal informant is not a fruitful one and that discrepancies do not reflect measurement error (De Los Reyes, 2011). Thus, the field has moved away from viewing discrepancies as reflecting error, with the consequence that these discrepancies should be minimized. Instead,
it is now believed that discrepancies themselves have meaning (De Los Reyes, 2011; De Los Reyes & Kazdin, 2005), such that discrepancies can be informative of severity of problems and treatment outcomes. Indeed, it has been found that discrepancies are associated with negative outcomes such as more risk taking, more psychopathology and less involvement of parents in intervention (e.g., De Los Reyes, Goodman, Kliever, & Reid-Quiones, 2010; Israel, Thomsen, Langeveld, & Stormark, 2007).

Still, what should researchers do when their aim is not to investigate discrepancies themselves, but when they do have multi-informant data? It has become clear that it is not a question of choosing the optimal informant, as this optimal informant does not exist. It is advised that researchers do take the overwhelmingly consistent findings on informant discrepancies into account, first, by hypothesizing a priori whether informants’ reports will converge or diverge, when they diverge what this may mean and when drawing conclusions from a study to be informant-specific (De Los Reyes, Thomas, Goodman, & Kundey, 2013). Essentially, this means that we cannot generalize findings from this thesis regarding problem behaviours to other informants than mothers, as we used only one informant for our dependent variable in all chapters. Stated otherwise, our findings should be interpreted as informant-specific. This means we may expect different findings when investigating the same construct using a different informant. Second, informant reports are not considered random, but influenced by attributions and biases of the behaviour being reported and the context this behaviour takes place in (De Los Reyes & Kazdin, 2005). Because these attributions, biases and contexts have been found to be associated with informant discrepancies, it seems logical that these factors should be taken into account when designing a new study. As such, researchers are advised to gather information about the constructs being measured, to study how observable the constructs under study are, in what context the constructs are being measured and what is already known about informant reports and discrepancies on these constructs. These precautions do not minimize informant discrepancies, but make these increasingly interpretable. Also, it is advised that due to the problems with informant reports, it may be better to use more measures of the same construct, instead of including many constructs in a study. By measuring what you really want to know very carefully, construct validity of the construct is increased (De Los Reyes, 2011).

Part I and II: Concluding statement

In the present thesis, we provided insight into the important role of both psychometrics and contextual factors in both parent, teacher, and child report of problem behaviours. The presented findings revealed that a more appropriate indicator of reliability, McDonald’s omega, yields more reliable indices for the SDQ’s subscales than the commonly used Cronbach’s alpha. Also, the findings highlighted the adequate five-factor structure of the SDQ and its robustness in terms of measurement invariance. Policy makers are advised to implement the SDQ as a screening instrument, to provide end-users with easily accessible and good-quality information on psychometrics of this instrument, and to open the debate concerning ethics in screening. As for contextual factors, findings revealed that parenting stress was positively related to the course of both in- and externalizing problems. Also, findings showed that children are similar to friends with internalizing problems, as reported by teachers. Given the relative scarcity of research into internalizing problems in early childhood, researchers are advised to increase research on this topic, thereby incorporating a holistic approach. Regarding child self-report of problem behaviors, findings highlighted the adequate stability and criterion validity of the Berkeley Puppet Interview. Still, inconclusiveness regarding reliability and construct validity of this age appropriate instrument warrant further research into psychometrics of this measure. As for specific parenting tactics, child perceived psychologically controlling parenting was robustly related to child reported in- and externalizing problems concurrently. Such a robust association could not be established for the link between maternal and child separation anxiety, and the mediating role of psychologically controlling parenting in this link. Again, these inconclusive findings warrant future research where child self-reports of both psychological control and child problem behaviours are combined with parent reports. In such studies, it is considered vital to take findings from the informant discrepancies literature into account.
References


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


Nederlandse Samenvatting
Publications
Dankwoord
Curriculum Vitae
Nederlandse Samenvatting

Probleemgedrag bij kinderen komt relatief vaak voor en heeft een veelheid aan negatieve consequenties voor de ontwikkeling van het kind zelf, maar ook voor ouders en de bredere sociale omgeving. In dit proefschrift wordt de rol van de omgeving in de samenhang met dit probleemgedrag belicht. Hierbij wordt speciale aandacht gegeven aan hoe dit probleemgedrag het beste te meten. Specifiek richtte dit onderzoek zich op twee thema’s. Het eerste thema heeft betrekking op de ouder- en leerkrachtrapportage van probleemgedrag bij jonge kinderen. Centraal hierin staat de vraag: wat zijn de psychometrische eigenschappen van een screeningsinstrument voor psychopathologie bij jonge kinderen? Daarnaast wordt de rol van omgevingsfactoren in de ontwikkeling van probleemgedrag bij jonge kinderen, zoals gerapporteerd door ouders en leerkrachten, onderzocht. Het tweede thema heeft betrekking op de zelfrapportage van jonge kinderen van hun probleemgedrag. Hierbij is de volgende vraag belangrijk: hoe zijn de psychometrische eigenschappen van een leeftijdsadegquaat instrument voor het beoordelen van percepties die kinderen hebben van hun probleemgedrag en hun ouders’ opvoedstijlen? Daarnaast wordt de rol van specifieke opvoeddimensies in relatie tot probleemgedrag bij jonge kinderen onderzocht.

Deel I. Probleemgedrag bij jonge kinderen: psychometrische en contextuele factoren

In Hoofdstuk 2 werd een overzichtsstudie gepresenteerd van studies naar de psychometrische eigenschappen van de SDQ. Kort gesteld behelst psychometrie of je meet wat je wilt meten (validiteit) en of deze metingen over tijd een vergelijkbaar zijn (betrouwbaarheid). Uit deze studie werd duidelijk dat de betrouwbaarheid van de subschalen van de SDQ, zoals gerapporteerd middels Cronbach’s alpha, niet altijd voldoende was, met name voor de ouderversie. Daarnaast werd gevonden dat de predictieve validiteit van de SDQ, in hoeverre de score op de SDQ over tijd voorspelling is voor bijvoorbeeld meer gebruik van zorg, zeer weinig is onderzocht. Vervolgens werd in eigen onderzoek bekeken of deze resultaten wellicht te verbeteren waren. Hiertoe werd de SDQ ingevuld door een groot aantal ouders uit een steekproef van kinderen van de basisschoolleeftijd. In deze studie, gepresenteerd in Hoofdstuk 3, werd gevonden dat de SDQ subschalen voldoende betrouwbaar zijn wanneer er een andere coëfficiënt voor het bepalen van de betrouwbaarheid wordt gehanteerd, namelijk coefficient Omega in plaats van Cronbach’s alpha. Vervolgens werd de originele vijffactor structuur, emotionele problemen, gedragsproblemen, hyperactiviteit-inattentie, problemen met leeftijdgenoten en prosociaal gedrag, gerealiseerd en werd meetinvariantie onderzocht. Meetinvariantie is een statistische maat die een indicatie geeft in hoeverre de betekenis van vragen voor verschillende groepen.

Na het onderzoeken van deze psychometrische eigenschappen richtten we ons op de rol van contextuele factoren in de verklaring van probleemgedrag bij jonge kinderen. In Hoofdstuk 5 bekeken we de overlap tussen internaliserende en externaliserende problematiek, en onderzochten we in hoeverre deze over tijd met elkaar samenhangen. We vonden dat er een sterke overlap is op één meetmoment, maar dat deze problematieken over tijd niet met elkaar samenhangen, wanneer gecorrigeerd wordt voor al bestaande niveaus van deze problematieken. Echter, wanneer het niveau van de problematieken werd meegenomen (hoe ernstig de problematiek is), vonden we dat de overeenkomst van problematieken werd doorgemaakt en dat de problematiek in de verdere ontwikkeling bij jonge kinderen Spinner speelde. We vonden dat jonge kinderen die een grotere mate van internaliserende problematiek hebben een grotere risico hebben om problemen te ontwikkelen. Deze resultaten geven een indicatie van de complexiteit die speelt in de interactie tussen ouder- en kindfactoren, en hoe deze complexiteit de ontwikkeling van probleemgedrag kan beschrijven.

De resultaten van dit proefschrift geven aanleiding tot het breder gebruiken van coëfficiënt omega als aanvulling op coëfficiënt alpha. Aan beleidsmakers wordt geadviseerd de SDQ te implementeren als screeningsinstrument, om gebruikers te voorzien van informatie over de psychometrische kwaliteiten van dit instrument en om het debat te openen over de ethiek van screening. Wat betreft contextuele factoren werd aan onderzoekers geadviseerd om meer onderzoek te doen naar internaliserende problemen, specifiek bij kinderen op jonge leeftijd, aangezien wordt aangenomen dat de voorwaarden voor het ontwikkelen van internaliserende problemen zich op jonge leeftijd ontwikkelen. Daarnaast wordt geadviseerd dat onderzoekers hierbij een holistische houding aannemen, zodat de versnappede kennis die op dit moment is over internaliserende problemen wordt geïntegreerd. Deze integratie zou ertoe kunnen leiden dat er een beter beeld ontstaat van de risicofactoren die er zijn voor het ontwikkelen van internaliserende problemen, waardoor kinderen en hun ouders beter kunnen worden geholpen om deze problemen te doen afnemen.

Deel II. Percepties van kinderen van probleemgedrag en opvoedstijlen

In Hoofdstuk 8 werden kinderen geïnterviewd door middel van een leeftijds-adequaat instrument, genaamd Berkeley Puppet Interview (BPI). In deze studie werd bekeken in hoeverre de antwoorden die de kinderen gaven betrouwbaar en valide waren. Van de stabiliteit en criterium validiteit van BPI scores zijn voldoende, met name voor de internaliserende schaal, en in mindere mate voor de internaliserende schaal. Als gevolg van de bi-modale verdeling van BPI scores is het complex om betrouwbaarheid en validiteit goed te onderzoeken. Dit punt verdient in toekomstig onderzoek aandacht, zodat bekeken kan worden in hoeverre het BPI betrouwbaar en valide is. Vervolgens werd in Hoofdstuk 9 bekeken in hoeverre een psychologisch controlerende opvoedstijl geassocieerd was met internaliserende en externaliserende problematieken, zoals gerapporteerd door kinderen. We vonden een sterke associatie tussen psychologisch controlerend opvoeden en probleemgedrag bij kinderen, zelfs wanneer we controleerden voor belangrijke opvoedomstandigheden zoals gedragsmatige controle en responsiviteit. Deze resultaten betekenen enerzijds dat de perceptie van kinderen belangrijk is in het bepalen van welke factoren samenhangen met probleemgedrag, en anderzijds dat psychologisch controlerend ouderschap sterk samenhangt met probleemgedrag. Tenslotte werd in Hoofdstuk 10 bekeken in hoeverre deze
psychologisch controlerende opvoedstijl ook een verklaring zou kunnen zijn voor verbanden tussen probleemgedrag bij moeders en hun kinderen. Meer specifiek bekeken we in hoeverre separatie angst bij moeders geassocieerd was met separatie angst bij kinderen, en of een psychologisch controlerende opvoedstijl deze relatie kon verklaren. We vonden dat er wel een samenhang was tussen separatie angst bij moeder en kinderen, maar dat deze relatie niet verklaard kon worden door een psychologisch controlerende opvoedstijl. Hoewel we vonden dat separatie angst bij moeder gerelateerd was aan separatie angst bij kinderen op één moment, bleek deze relatie geen stand te houden over tijd.

Gezien er naar de rol van psychologisch controlerend opvoeden bij jonge kinderen weinig onderzoek gedaan is, en uit dit proefschrift enkele bevindingen naar voren komen die nog niet geheel uitsluitsel geven over de rol van deze opvoeddimensie, wordt het belangrijk gevonden om in de toekomst meer onderzoek te doen naar de relatie tussen psychologisch controlerend opvoeden en de ontwikkeling van probleemgedrag. In dit onderzoek zou een combinatie gemaakt moeten worden van kind- en ouderrapportages om meer duidelijkheid te geven over de perceptie van beide informanten. Voor het BPI geldt ook dat er meer onderzoek nodig is voor dat dit instrument geïmplementeerd zou kunnen worden, waarbij met name onderzoek naar de psychometrische eigenschappen en de toepassing voor de klinische praktijk nodig is.

Publications

This Thesis


Other Publications


Acknowledgements

This work was granted by the Dutch Organization for Health Research and Care Innovation (ZonMw) [grant number: 157002001].

Dankwoord

Eindelijk, het is af! Graag richt ik me hier tot iedereen die geholpen heeft dit proefschrift tot stand te brengen.

Jan, dank voor je vertrouwen in mij en je immer scherpe ogen voor zo'n beetje alle situaties die we tegenkwamen in het project: van ingewikkelde statistische problemen tot tekstuele onvolkomenheden. Dank ook dat je me hebt laten zien dat het steeds belangrijk is oog te houden voor menselijke verhoudingen, op die manier zoals alleen jij dat kan, met een altijd vernieuwende mix van woorden waar de bonden vloeken zo tegen is. Rutger, dank dat jij in mij geloofde en me stimuleerde om meer te doen dan ik misschien zelf dacht dat ik kon. En zie daar, het is gelukt. Die ervaringen motiveren mij nu nog steeds om mijn ambities na te blijven jagen. Nu je niet meer dagelijks op de afdeling te vinden bent merk ik ook hoeveel energie en levendigheid je bracht; dank dat ik daarvan heb kunnen profiteren en ook daarom zo’n mooie tijd heb gehad als ajo. Roy, dank dat jij me alle ruimte gaf om het project zo in te vullen als ik dacht dat goed was, maar ook dat je er altijd was om me gerust te stellen als de stress te veel opliep. Je was letterlijk de veilige haven (al zal je dit zelf denk ik te zweerig vinden); hoe je het ook wil noemen, je bent voor mij een hele fijne dagelijks begeleider geweest. Ik waardeer onze gesprekken over werk en niet-werk en je kantoorhumor, waaronder de pratende koffiepot op de werkwk in het bijzonder.

Ad, zonder jouw hulp en vooral je geduld met het uitleggen van statistiek was het niet zover dat je het scheepje kon laten zinken van zo’n betekenisvolle voorstelling. Je had me laten zien dat het zo belangrijk is altijd terug te gaan naar de basis wanneer je er niet uit komt. Bart, dank je voor je mederenden en – schrijven bij onze artikelen over psychologisch controlerend opvoeden; ik heb je theoretische inbreng en actieve co-auteurschap erg gewaardeerd. Emmanuell, thank you for your help with the statistical analyses for the BPI data and another thank you for your warm hospitality during my stay in Lausanne. Mara Brendgen, without your input, the homophily paper wouldn’t be as good as it turned out to be, thank you. Jeffrey Measelle, thank you for training me to become a Berkeley Puppet Interview puppeteer, I had great fun learning the skills.

Dan, iedereen die ‘op de vloer’ heeft meegewerkt aan de noeste arbeid van scholen zoeken, ouders aanschrijven, leerkrachten bellen, kinderen interviewen, vragenlijsten checken en data invoeren. Josje, Esther, Laurien, Jellie, Funda en Maaike dank jullie zorgvuldige inzet voor het project en jullie bereidheid heel Nederland te bereizen om al die kinderen te interviewen. De studenten die hun scriptie hebben geschreven en ook heel veel data hebben verzameld voor het project; dank jullie wel Vera, Laura, Ratna, Birgit, Ellen, Esther, Renske, Lian, Nicole, Sharon en Susanne. Wendy en
Marijke, jullie inzet heeft ervoor gezorgd dat ik in het jaar waarin we naast het lopende project ook nog eens kinderen gingen interviewen met het poppeninterview niet gillend gek ben geworden. Dank dat jullie me zoveel werk uit handen hebben genomen, en dat ik dit jullie kon toevertrouwen. De puppeteers, dank voor jullie geweldige inzet en jullie vermogen om een groot deel van jullie dag door te brengen als Iggy en Ziggy zonder een identiteitscrisis te ontwikkelen: Wendy, Marijke, Birgit, Laura, Marloes, Carlijn, Vivianne en Suzanne.

Collega’s van de 6e en het Ambu, jullie hebben misschien niet direct bijgedragen aan het project, maar toch zeker indirect met jullie gezelligheid op borrels, werkweken en congressen, geestverruimende praatjes na een ochtend schrijven en steun wanneer het even niet lukte met de zelfdiscipline; dank daarvoor! Ook de dames van het secretariaat: Diana, Kalja en Lon bedankt voor jullie gezelligheid en praktische hulp die voor de warige wetenschapper broodnodig is. Marieke, mijn roomie van het eerste uur, dank voor je hulp bij alle kwesties die je als kamergenoten tegenkomt in de vijf jaar dat je samen opgehoekt zit, en vooral voor de mooie en gezellige tijd die we samen hebben gehad, je bent een topster!renske, dank voor je imomvloijke aanwezigheid en je steun, Suus dank voor je oprechtheid, belangstelling en je humor, Manon dank voor je optimisme en je raad en daad waarvoor ik altijd bij je terecht kan, Yuli, dank voor je luisterend oor en het feit dat je me van mijn computer vandaan trekt voor een wandeling. Met jullie was het altijd gezellig en dat heeft mijn aio-tijd in de moeilijke, maar gelukkig ook in de niet moeilijk tijden, een stuk leuker gemaakt! Marleen en Monique, fijn dat ik bij jullie altijd kon komen aanwaaien voor een praatje en wat grappen en grollen. Carmen, wat fijn dat we om die trilogieën heen zo leuk kunnen praten over het leven, en gelukkig ook over alle minder hoogdragende zaken die op een afdeling de ronde kunnen doen. Rowella, wij waren toch altijd twee rare eenden in de bijt op de 6e met onze psychometrie projecten. Fijn dat we daar altijd over konden sparren en elkaar konden helpen met prangende (statistiek)vragen. Dank dat je altijd op scherp staat waardoor je me helpt om kritisch te blijven en meer uit mezelf te halen, dat je me laat zien dat het mogelijk is klinisch werk te combineren met onderzoek én dat je altijd tijd maakt voor me waarin je me helpt met rake adviezen.

Lieve vrienden, dank dat jullie me hebben geholpen dit proefschrift af te ronden door ervoor te zorgen dat ik het werk ’s avonds en in het weekend met een gerust hart kon laten voor wat het was omdat ik leuke dingen ging doen met jullie. Lieve vriendinnen uit het Brabantse land, Rie, Mo, Fran, Ils, Mar en Inge; wat vind ik het geweldig dat we al zo lang bevriend zijn! Die lange tijd maakt dat we zo’n beetje alles van elkaar kennen en dat waarderen ik enorm. Samen hebben we nog steeds de grootste lol en ik hoop dat dat nog lang zo blijft. Lieve Loes, dank dat de deur bij jou altijd open staat voor mij, je warmte en je luisterend oor. Naast onze serieuze gesprekken geniet ik van de momenten waarbij we samen in een lachstuip terecht komen die niet veel mensen om ons heen lijken te begrijpen, wat de lol voor ons alleen maar groter maakt. Lieve Chris, dank voor je energie en verfrissende kijk op het leven, waardoor je mij ook steeds weer anders naar het leven laat kijken. Zonder de ‘input’ die jij nodig hebt waren we vast niet op al die mooie feesten en wie weet wat nog meer terecht gekomen waar we zoveel lol hebben gemaakt. Loes en Chris, dank dat jullie zulke lieve en fijne vriendinnen zijn! Tobias, Guus, Jeroen en Renee, Pieter en Desiree, Wouter en Kimberly, Luke, Paul, dank dat jullie zulke fijne vrienden zijn, dat we altijd samen een biertje kunnen drinken en voor alle leuke dingen die we samen hebben meegemaakt sinds groep P. Frau Bakker en Herr N, dank voor de grootste lol die we samen hebben over het wel en wee van de wetenschappelijke wereld en daarbuiten, onze avonden samen zijn me dierbaar! Lieve Marjolein, wat fijn dat we elkaar sinds Curacao nog weten te vinden; ik vind ons een kei goeie mix van Twentse en Brabantse gezelligheid; da is. Dennis en Brenda, dank dat jullie zo attent zijn in alles wat jullie doen, dat maakt dat ik me erg welkom voel bij jullie en in de vriendengroep.

Lieve Nettie en Coen, dank voor jullie enthousiaste interesse in mijn werk, maar vooral voor jullie warmte en steun. Hoe jullie mij laten zien dat je je dromen kan verwezenlijken door dit zelf te doen vind ik geweldig, en geeft me vertrouwen dat dat ook zal lukken. Lieve Maart, dank dat je mij in onze gesprekken altijd laat kijken naar hoe het óók kan en voor alle fijne tijden die we samen hebben beleefd op de Museum Kam. Ik denk daar met erg warme herinneringen aan terug. Dear family, thank you for being there: Mam, Pap, Lara, David, Alfred, Carmena, Paul, Joshua, Michael, James, Alicia, Liz and Roman. Lieve Laar, onze levens zijn zo anders, maar toch weten we elkaar altijd weer te vinden in onze ruchtere kijk op het leven, en hebben we de grootste lol ook al hebben we elkaar een tijd niet gezien. Dank dat je gewoon de lieve Laar bent gebleven die je altijd al was, ook al ligt de halve wereld aan je voeten (toch best een opgave!). Dear Mena, thank you for being such a great big sister. How you manage to be so attentive while also making a career and running a family is a mystery to me; in any case, you are a great example to me. Lieve Pap en Mam, dank voor het warme nest wat jullie me gegeven hebben en jullie steun en vertrouwen in alles wat ik doe. Lieve Mam, dank je dat je mij hebt geleerd nuchter en voor alle fijne tijden die we samen hebben beleefd op de Museum Kam. Dank dat je gewoon de lieve Laar bent gebleven die je altijd al was, ook al ligt de halve wereld aan je voeten (toch best een opgave!). Lieve Papa en Mam, dank voor het warme nest wat jullie me gegeven hebben en jullie steun en vertrouwen in alles wat ik doe. Lieve Laar, dank je dat je mij hebt geleerd nuchter en voor de grootste lol die je altijd al was, ook al ligt de halve wereld aan je voeten (toch best een opgave!).
Curriculum Vitae

Lisanne Stone was born May 10th 1985 in Geldrop, the Netherlands and grew up in neighboring village Mierlo. After completing secondary education at Strabrecht College in Geldrop in 2003 she moved to Nijmegen to study Clinical and Developmental Psychology at Radboud University Nijmegen. She spent six months in Curaçao for a clinical internship and obtained her master’s degree in Developmental Psychology in 2008. In December 2008, she started as a PhD student at the Behavioural Science Institute, Radboud University Nijmegen. She was responsible for the organization and execution of a large multi-wave multi-informant study on the psychometric properties of a screening instrument for psychopathology in young children. In addition to this study, she set up a smaller study wherein she was interested in investigating psychometrics of a child self-report instrument; the Berkeley Puppet Interview. To this end, she collaborated with Prof. Dr. Jeffrey Measelle and became a certified Berkeley Puppet Interview trainer and interviewer. During several work visits, she benefited of the guidance and knowledge of several scholars; first, she visited Prof. Dr. Mara Brendgen in Canada with whom she collaborated on her paper concerning friendships of young children. Next, she visited Prof. Dr. Bart Soenens in Belgium to collaborate on her work regarding child perceptions of parental psychological control, and finally she visited Dr. Emmanuel Kuntsche in Switzerland where she learned new data analysis techniques. Further, she presented her work at several national and international conferences and supervised numerous master’s theses. Importantly, she always combined her research with clinical work. Specifically, she worked as a clinical child psychologist at an outpatient clinic (Ambulatorium Nijmegen) during her PhD-project. She strongly believes the collaboration between research and practice is imperative for the development of these two fields. Therefore, she started her postdoctoral training to become a registered mental health psychologist (GZ-psycholoog) in 2013 at the Ambulatorium. Currently, she works a teacher at Radboud University Nijmegen where she coordinates and teaches a course on advanced clinical skills and as a child psychologist at Ambulatorium Nijmegen. Shortly, she will continue her clinical work at ProPersona where she will finish the final two years of her training as a child psychologist. In the future, she aspires to bring her research skills to practice-driven research, as well as to continue her research on psychometrics, and finally to combine this with clinical work.

mooie dingen ondernemen (met dank aan jouw haast onuitputtelijke levenslust en energie), lol maken en genieten van het goede leven dat we samen hebben!