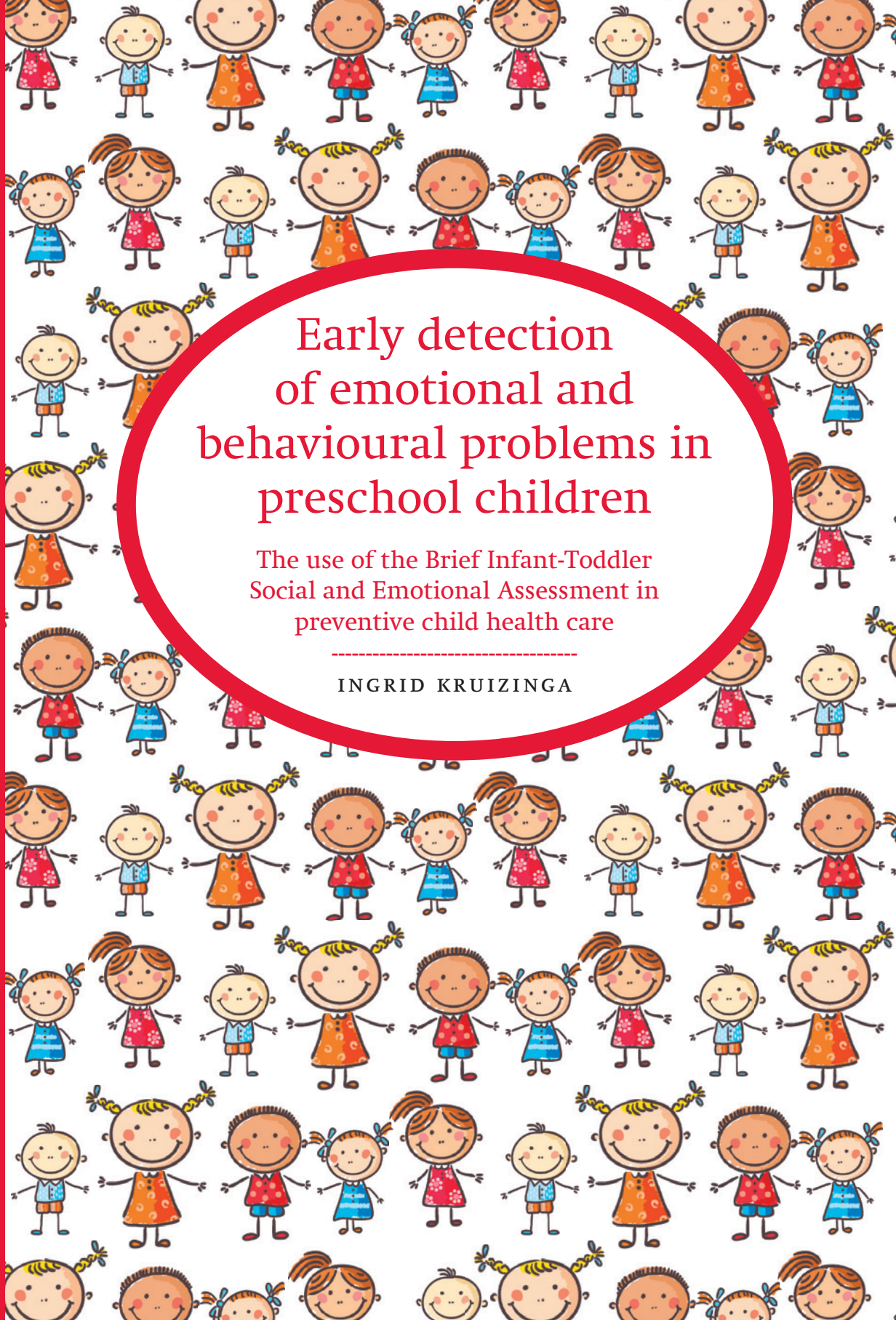


Early detection of emotional and behavioural problems in preschool children

The use of the Brief Infant-Toddler Social and Emotional Assessment in preventive child health care

INGRID KRUIZINGA

Early detection of emotional and behavioural problems in preschool children

The use of the Brief Infant-Toddler Social and Emotional Assessment in preventive child health care

---

INGRID KRUIZINGA

Acknowledgements

We gratefully acknowledge the contribution of the preventive child health care organizations and their professionals in the larger Rotterdam area to this study.

The work presented in this thesis was conducted at Erasmus Medical Centre, department of Public Health, and the Municipality of Rotterdam, and supported by a grant from ZonMw (Grant Number 80-82435-98-8058).

Cover	Witte Rook
Layout	Renate Siebes, Proefschrift.nu
Printed by	Ridderprint, Ridderkerk
ISBN	978-94-90791-32-2

Early Detection of Emotional and Behavioural Problems in Preschool Children

The use of the Brief Infant-Toddler Social and Emotional Assessment in preventive child health care

Vroegsignalering van emotionele en gedragsproblemen bij peuters

De toepassing van de Brief Infant-Toddler Social and Emotional Assessment in de preventieve jeugdgezondheidszorg

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
woensdag 11 maart 2015 om 15:30 uur

door

Ingrid Kruizinga

geboren te Voorburg

Promotiecommissie

Promotor	Prof.dr. H. Raat
Overige leden	Prof.dr. F.C. Verhulst
	Prof.dr. B.J.C. Middelkoop
	Prof.dr. L.R. Arends
Copromotor	Dr.ir. W. Jansen

Paranimfen	Denise van Bemmel, MSc.
	Dingenita Leonie van Dijk, MSc.

Contents

Chapter 1

General introduction

The main aim of this thesis was to evaluate the role of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) as an instrument for the early detection of psychosocial problems in toddlers, in the setting of preventive child health care. More specifically, the psychometric properties and screening accuracy were evaluated. Additionally, in a cluster randomized trial, we assessed whether the health outcomes (specifically psychosocial well-being) were better in the intervention group, where the BITSEA was applied to detect psychosocial problems, compared to the outcomes in the control group where 'care as usual' (see below) was applied. The follow-up duration was one year. In this general introduction the prevalence of psychosocial problems in toddlers, its impact and early detection will be addressed. The chapter finishes with a summary of the research questions and an outline of the thesis.

Psychosocial problems

The prevalence of psychosocial problems is already relatively high in preschool children. The prevalence rates differ between studies, depending on the measuring method that is used. In the Netherlands, 5-10% of the preschool children score in the clinical range on a questionnaire that measures psychosocial problems, such as the Child Behavior Checklist (CBCL) [1] or the Infant-Toddler Social and Emotional Assessment (ITSEA) [2-5]. And child health professionals identify psychosocial problems in 7-25% of the preschool children [2-5]. Prevalence rates differ also between subgroups: boys experience psychosocial problems more often than girls [5,6] and, in the Netherlands, psychosocial problems are more often reported in non-native children compared to native children [7,8].

Psychosocial problems is a collective term for a great variety of problems. Often these problems are divided in two components; externalizing and internalizing problems.

- Externalizing problems include problems such as; aggressive behaviour or overactivity.
- Internalizing problems include problems such as; anxiety, depression or withdrawal.

Additionally, another component of psychosocial problems is identified for problems that are not easily categorized under externalizing of internalizing problems [9]. Examples of such problems are sleeping or eating problems (i.e. 'dysregulation'). There is evidence that children with regulatory problems are vulnerable to developing social-emotional or behavioural

problems and psychiatric disorders [10]. This suggests that difficulties with regulation may play a role in the emergence or maintenance of psychopathology.

The reason for child health professionals to refer a child changes with the child's age: in the first year children are most often referred because of eating and sleeping problems, in their second and third year mostly for eating and behavioural problems [11]. An important question regarding such a referral decision is; when are behaviour and/or certain emotions considered a problem? Many children are sometimes anxious, but this is not always considered a problem and it does not always require a diagnosis and treatment. A good indicator for the significance of problems is parental concern about the child's development [12,13].

A disorder that is one of the earliest to raise concerns by parents, and one of the earliest to be diagnosed in a child, is autism spectrum disorder (ASD). An onset of symptoms prior to 36 months is a mandatory criterion for diagnosis in both ICD-10 [14] and DSM-IV [15]. However, ASD is rarely diagnosed before the age of 36 months. Studies by De Giacomo and colleagues [16] and Howlin and colleagues [17] have shown that the average age of the child when parents became concerned with the development was around 19 months. The average age of the child at which parents first sought help was around 2 years [16,17] and the average age of final diagnosis was around 5.5 years [17]. ASD and psychosocial problems can have a very negative impact on development and persevere later in life. Research has demonstrated that problems at preschool age are relatively stable over time [18-20]. Briggs-Gowan and colleagues [18] found that children with a high score on an early detection instrument at age 12-36 months still had internalizing and externalizing problems at age 6 (as indicated by the CBCL). According to a study of Mesman and colleagues [20], a child of 11 years old has 3 and 5 times more chance of still having respectively an internalizing or externalizing disorder (as described in the DSM-IV), when the child had internalizing or externalizing problems (as indicated by the CBCL) at age 2- to 3-years. Moreover, Lavigne and colleagues [21] showed that more than 50% of children with psychosocial problems at age 2- to 3-years (as indicated by the CBCL), continued to have some psychiatric disorder (as described in the DSM-III) 3 to 4 years later.

It has been recommended that psychosocial problems should be detected at a very young age and followed by appropriate care [22,23]. Measurement, early detection and treatment of psychosocial problems at a young age are important because this may contribute to a reduction of problems and an increase of competencies at older ages [24,25]. For instance, Elliot and colleagues [25] showed that a low-intensity intervention programme, delivered in a community setting targeted at preschool children, can have positive effects in reducing

major behavioural problems in early childhood, namely hyper-active/distractable behaviour, 2 years after the completion of the programme. Also, Reynolds and colleagues [26] found that participation in an established early childhood intervention was associated with better educational and social outcomes up to the age of 20 years. In a professional guideline for Youth Health Care in the Netherlands the usefulness of early detection instruments and subsequent interventions is acknowledged [27]. In this guideline it is recommended to use questionnaires to recognize psychosocial problems, to prevent them or reduce their impact on both the child's life and society. In a paper by Heckman [28], the impact of interventions on society is expressed in an estimated 'rate of return to investment'. Assuming the same investment at each age, the rate of return to investment while a person is young is estimated to be higher than the rate of return at later age: The positive effects of early investments are beneficial over a longer period of time. In addition, cognitive skills acquired early on may facilitate later learning and functioning, leading to cumulative benefit. See Figure 1.1.

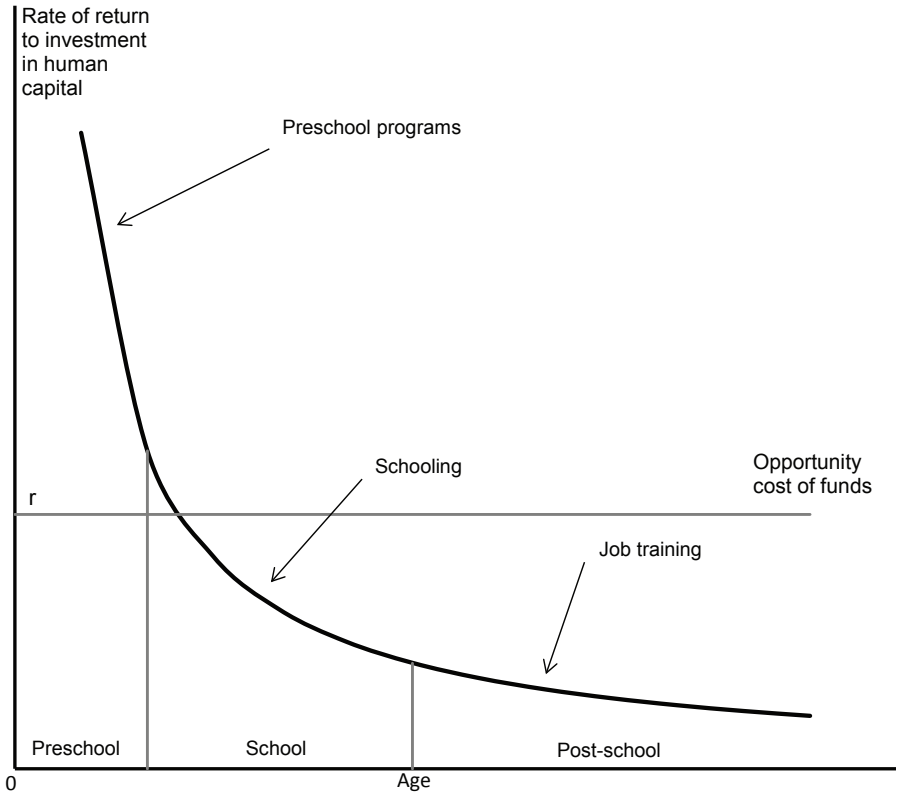


Figure 1.1 Rates of return to human capital investment, assuming investment is equal across all ages (Heckman, 2008, p.311. Printed with permission from the author.).

Dutch preventive child health care and early detection

In the Netherlands, the health care system offers publicly funded preventive programs for all children from birth to the age of 19 years. As a part of this system, parents of children aged 0- to 19-years old are invited to routine primary care visits (i.e. well-child visits). To ensure that each child in the Netherlands is offered the same minimum level of services, basic services of preventive youth health care are defined in the Basic Working Package for Health Care in 0-19 year olds (*Basistakenpakket Jeugdgezondheidszorg 0-19 jaar*) [29]. This setting offers a great opportunity for the early detection of psychosocial problems.

Early detection of psychosocial problems is an approach for sorting those children who probably have problems from those who probably do not. The group with probable difficulties are typically referred for diagnostic evaluations and if diagnosed, referred for treatment [30]. This is a simplified depiction of the process, in reality there are many factors playing a role in the pathway from problem recognition to problem treatment. Godoy and Carter [31] proposed a model of the influences on parent help seeking behaviour and child health care provider recommendations, see Figure 1.2.

Several aspects of this model are incorporated in this thesis: Prior to working with the BITSEA, child health professionals were *trained* by the researchers with support of a specialized psychiatrist in the use and scoring of the BITSEA. The child health professionals made their *assessment* of the development of the child based on *completed screening forms*, the *perceptions of child behaviour*, *worry about the child's behaviour by the parent*, *prior knowledge of the family* and *parent-provider conversation*. According to this assessment the child health professional *decides to recommend help* or not. In our study, the (pursuit of) referral was monitored one year after the well-child visit, in order to evaluate whether parent's *sought out the recommended help*. Special attention was paid to *child characteristics* by evaluating subgroups of child gender and ethnicity. In this model, the early detection instrument operates as a catalyst on 1) the parent/provider appraisal processes and 2) help-seeking/recommendation/change spectrum. Theunissen and colleagues [32] evaluated the merit of the use of an early detection instrument above the sole 'impression' of a child health professional. They found that child health professionals were not in all experimental groups very accurate in identifying children with psychosocial problems. However, when an early detection instrument was used, the identification of children with psychosocial problems improved significantly. These results imply that the use of an early detection instrument is effective in correctly identifying children with psychosocial problems.

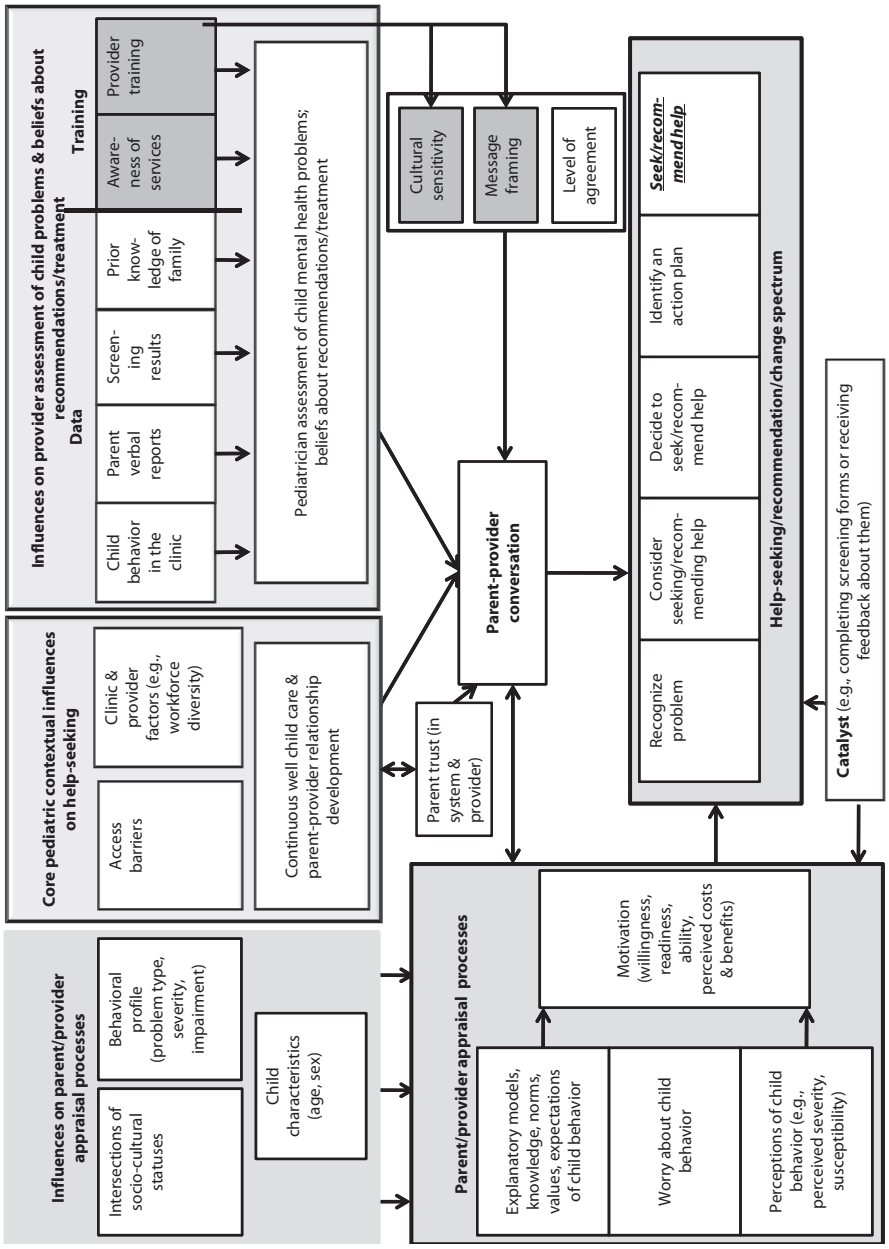


Figure 1.2 Influences of parent help seeking and child health care provider recommendations (Godoy, L. and A.S. Carter (2013), p. 83. Printed with permission from the authors).

Although, as mentioned before, early detection of psychosocial problems improves the prognosis of children, several studies indicate that improvement in the quality of early detection is still warranted. Reijneveld and colleagues [4] found that child health professionals did not identify psychosocial problems in 71% of the children with psychosocial problems (i.e. a score in the clinical range of the CBCL, of note is that the child health professionals were blind for this score). In the study of Brugman and colleagues [2], this percentage was 43.

Early detection

Early detection instruments for psychosocial problems, intended for use in the preventive child health care, should have adequate psychometric properties, and should also be short, easy to administer, score and interpret [30]. Furthermore, early detection instruments should be able to correctly discriminate children with and without psychosocial problems. Of course the process of early detection will not be without errors, but should be as accurate as possible to minimize the expenses associated with over-referrals and under-detection [30]. Early detection instruments should be evaluated in a large sample, whose characteristics reflect those of the total general population in terms of, among others, ethnicity and gender [30].

In the setting of preventive child health care, general early detection instruments are warranted, since the aim is to early detect a broad range of possible psychosocial problems. Short comprehensive instruments that are appropriate to measure psychosocial problems in children of preschool age are limited [33]. The Child Behavior Checklist 1.5-5 (CBCL1.5-5) and Infant-Toddler Social and Emotional Assessment (ITSEA) are early detection instruments that are well-validated and measure a broad range of psychosocial problems, and in the case of the ITSEA also delays in competencies. However both instruments are too extensive to apply in the context of well-child visits. Short questionnaires, such as the Eyberg Child Behavior Inventory [34] or the Toddler Behavior Screening Inventory [35], only measure problem behaviour and do not address social-emotional competencies. Measuring delays in social-emotional competence, however, is also important since delays in the acquisition of competencies are strongly related to a wide range of psychosocial problems later in life [36], and are often the prodromal signs of developmental disorders, such as ASD [37].

Widely used in the Netherlands, for the early detection of psychosocial problems in preschool children, is an questionnaire specifically developed for the use in the Dutch preventive child health care: the KIPPPI (KIPPPI is a Dutch acronym for Brief Instrument Psychological and

Pedagogical Problem Inventory) [38]. The questionnaire measures psychosocial problems in 2-year olds, that might be possible pedagogical challenges for the parents. The KIPPPI has 67 items and consists of a Wellbeing scale (31 items), Competence scale (25 items) and an Autonomy scale (11 items). As many aspects of psychosocial development are addressed in the questionnaire, the KIPPPI can be used by the child health professional to guide conversation with the parent. However, the KIPPPI is relatively long and the scientific base for scoring and norming the answers is lacking.

One questionnaire that was recommended by Hermanns and colleagues [39] for the purpose of early detecting psychosocial problems in toddlers is the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) [40]. The BITSEA is a short instrument that consists of 42 items and is comprised of two scales, a Problem scale (31 items) and Competence scale (11 items). Responses can be summed for each scale. The Problem scale assesses social-emotional/ behavioural problems such as aggression, defiance, overactivity, negative emotionality, anxiety, and withdrawal; a high score is less favourable. The Competence scale assesses social-emotional abilities such as empathy, prosocial behaviours, and compliance; a low score is less favourable [41]. The BITSEA also consists of 17 items that are specifically included for the early detection of ASD, belonging to either the Problem scale or the Competence scale. Additionally, the BITSEA also consists of 14 items that describe behaviours that may be indicative of a clinically significant problem even in the absence of a Problem or Competence score in the clinical range. Several of these items describe behaviour that may endanger a child (e.g. *'Does not react when hurt'*). The items intended to measure ASD or potential dangerous behaviour formally do not represent separate scales. Furthermore, the BITSEA has two single-item questions on parent worries regarding child language development and child behaviour, emotions or relationships. Studies on the BITSEA in the USA, Finland, Turkey and France have reported acceptable psychometric properties (i.e. reliability and validity) [40,42-45]. However, the BITSEA was not yet evaluated for the Dutch population and this was the focus of the study presented in this thesis.

Aim and study questions addressed in this thesis

The central aim of this study is to evaluate the reliability and validity of the Dutch version of the BITSEA in 2-year old children. Also, in a cluster randomized trial, the aim was to assess whether health outcomes (specifically psychosocial well-being) were better in the intervention group where the BITSEA was applied to detect psychosocial problems compared to the outcomes in the control group where 'care as usual' was applied (i.e. use of the KIPPPI);

the follow-up duration was one year. We hypothesized that the psychosocial wellbeing one year after the well-child visit, is more favourable for children whose parents completed the BITSEA, compared to children who received care as usual, since a more appropriate referral of children with psychosocial problems with consequently more appropriate care is expected when the BITSEA is applied, compared to 'care as usual'.

In this study, the psychometric properties and effectiveness of the BITSEA were evaluated for the total population as well as for subgroups of child gender and ethnicity, because between these subgroups there might be possible different predispositions for experiencing psychosocial problems, or cultural differences (i.e. interpretation of behaviour or question items).

The following research questions are addressed:

1. What is the reliability and validity of the BITSEA for the total population as well as for subgroups of child gender and ethnicity? (chapter 2)
2. What is the reliability and validity of the KIPPPI? (chapter 3)
3. What are the screening accuracy and cutpoints of the BITSEA for the total population as well as for subgroups of child gender (chapter 4&5)
4. Is there a difference, a year after the well-child visit, in psychosocial wellbeing between children whose parents completed the BITSEA and children who received care as usual? (chapter 7)

Outline of this thesis

This thesis consists of two parts:

The first part describes the evaluation of the reliability (i.e. internal consistency, test-retest reliability and interrater reliability) and validity (i.e. concurrent validity and discriminative validity) of the BITSEA (chapter 2). The reliability (i.e. internal consistency, test-retest and reliability), validity (i.e. concurrent validity and discriminative validity), screening accuracy (i.e. calculating ROC-curves) and clinical application (i.e. the relation between scale scores and the clinical decision of the child health professional) of the KIPPPI (chapter 3). Additionally, the clinical application of the BITSEA was evaluated (chapter 4). Furthermore, the screening accuracy of the BITSEA was evaluated in comparison to two golden standards; a) a score in the clinical range of the CBCL (chapter 4), and b) an ASD diagnosis (chapter 5). In part 1 cross sectional data was used that was gathered at the baseline of the study in the general population. Also data in a subsample of referred children was used.

Table with an overview of the studies presented in this thesis

Chapter	Instrument	Sample	Subgroups[1]	N	Focus
		Part 1 - The psychometric properties of early detection instruments			
2	BITSEA	general population	gender and ethnicitcy	3127	reliability & validity
3	KIPPPI	general population	none	2732	reliability, validity, screening accuracy & clinical application
4	BITSEA	a) general population; b) children with a clinical score on the CBCL	gender and ethnicitcy	a) 2060 b) 43	screening accuracy and clinical application
5	BITSEA	a) general population; b) children with a diagnosis of ASD	gender	a) 3127 b) 159	screening accuracy
		Part 2 – The evaluation of the Brief Infant-Toddler Social and Emotional Assessment			
6	-	-	-	-	study protocol
7	BITSEA; KIPPPI	general population	gender and ethnicity	a) 1207 b) 1403	differences between intervention (BITSEA) and control group (KIPPPI) in CBCL score at 1-year follow-up

[1] Subgroups were examined for child gender (boys & girls) and for child ethnicity (native Dutch & non-native Dutch).

The second part of the thesis describes the design of a longitudinal randomized controlled trial with a baseline measurement and a one-year follow-up measurement in which the effectiveness of the use of the BITSEA as an early detection instrument was compared to care as usual (chapter 6); the results are presented in chapter 7.

REFERENCES

1. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington: VT: University of Vermont, Research Center for Children, Youth, and Families.
2. Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP (2001) Identification and management of psychosocial problems by preventive child health care. Archives of Pediatrics & Adolescent Medicine 155: 462-469.
3. Klein Velderman M, Crone MR, Wiefferink CH, Reijneveld SA (2010) Identification and management of psychosocial problems among toddlers by preventive child health care professionals. European Journal of Public Health 20: 332-338.
4. Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP (2004) Identification and management of psychosocial problems among toddlers in Dutch preventive child health care. Archives of Pediatrics & Adolescent Medicine 158: 811-817.
5. Zeijl E CM, Wiefferink K, Keuzenkamp S, Reijneveld M (SCP/TNO Kwaliteit van Leven: Den Haag/Leiden; 2005.) Kinderen in Nederland [Children in the Netherlands].
6. Kraemer S (2000) The fragile male. British Medical Journal 321:1609.
7. Reijneveld S, Harland P, Brugman E, Verhulst F, Verloove-Vanhorick S (2005) Psychosocial problems among immigrant and non-immigrant children - Ethnicity plays a role in their occurrence and identification. European Child and Adolescent Psychiatry 14: 145-152.
8. Vollebergh WAM, ten Have M, Dekovic M, Oosterwegel A, Pels T, et al. (2005) Mental health in immigrant children in the Netherlands. Social Psychiatry and Psychiatric Epidemiology 40: 489-496.
9. National Center for Toddlers and Families (1994), Diagnostic Classification: 0-3. Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood. Washington, DC: National Center for Toddlers and Families.
10. Rothbart MK, Ahadi SA (1994) Temperament and the development of personality. J Abnorm Psychol 103: 55-66.
11. Keren M, Feldman R, Tyano S (2001) Diagnoses and interactive patterns of infants referred to a community-based infant mental health clinic. Journal of the American Academy of Child and Adolescent Psychiatry 40: 27-35.
12. Glascoe FP (1997) Parents' concerns about children's development: Prescreening technique or screening test? Pediatrics 99: 522-528.
13. Ellingson KD, Briggs-Gowan MJ, Carter AS, Horwitz SM (2004) Parent identification of early emerging child behavior problems - Predictors of sharing parental concern with health providers. Archives of Pediatrics and Adolescent Medicine 158: 766-772.
14. World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: WHO.

15. American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

16. De Giacomo A, Fombonne E (1998) Parental recognition of developmental abnormalities in autism. European Child and Adolescent Psychiatry 7: 131-136.

17. Howlin P, Asgharian A (1999) The diagnosis of autism and Asperger syndrome: findings from a survey of 770 families. Developmental Medicine and Child Neurology 41: 834-839.

18. Briggs-Gowan MJ, Carter AS (2008) Social-emotional screening status in early childhood predicts elementary school outcomes. Pediatrics 121: 957-962.

19. Lavigne JV, Gibbons RD, Christoffel KK, Arend R, Rosenbaum D, et al. (1996) Prevalence rates and correlates of psychiatric disorders among preschool children. Journal of the American Academy of Child and Adolescent Psychiatry 35: 204-214.

20. Mesman J, Koot HM (2001) Early preschool predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry 40: 1029-1036.

21. Lavigne JV, Arend R, Rosenbaum D, Binns HJ, Christoffel KK, et al. (1998) Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry 37: 1246-1254.

22. American Academy of Pediatrics. Committee on children with disabilities: development surveillance and screening of infants and young children. (2001) Pediatrics 108: 192-196.

23. (2003) Horen, zien, niet zwijgen [See all, hear all, say something]. Inspectie voor de Gezondheidszorg, Inspectie van het Onderwijs, Inspectie Openbare Orde en Veiligheid, Inspectie Jeugdhulpverlening, Jeugdbescherming.

24. Durlak JA, Wells AM (1998) Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. American Journal of Community Psychology 26: 775-802.

25. Elliot J, Prior M, Merrigan C, Ballinger K (2002) Evaluation of a community intervention programme for preschool behaviour problems. Journal of Paediatrics and Child Health 38: 41-50.

26. Reynolds AJ, Temple JA, Robertson DL, Mann EA (2001) Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. JAMA 285: 2339-2346.

27. Postma S (2008) JGZ-richtlijn. Vroegsignalering van psychosociale problemen [Guideline early detection psychosocial problems in preventive child health care]. Bilthoven: RIVM, Centrum Jeugdgezondheidszorg.

28. Heckman JJ (2008) Schools, Skills, and Synapses. Econ Inq 46: 289.

29. Ministry of Health, Welfare and Sports (2002), Basistakenpakket jeugdgezondheidszorg 0-19 jaar [Basic Working Package for Health Care in 0-19 year olds]. The Hague: Ministry of Health, Welfare and Sports.

30. Glascoe FP (2005) Screening for developmental and behavioral problems. Mental Retardation and Developmental Disabilities Research Reviews 11: 173-179.

31. Godoy L, Carter AS (2013) Identifying and Addressing Mental Health Risks and Problems in Primary Care Pediatric Settings: A Model to Promote Developmental and Cultural Competence. American Journal of Orthopsychiatry 83: 73-88.

32. Theunissen M H C, De Wolff M S, Vogels A G G, A R S (2011) Vergelijking van de ASQ:SE, BITSEA, KIPPPI en SDQ als signaleringsinstrument voor problemen bij 0-4 jarigen in de Jeugdgezondheidszorg [Comparison of the ASQ:SE, BITSEA, KIPPPI and SDQ as early detection tools for problems in 0-4 year olds in preventive child health care]. Leiden: TNO.

33. Glascoe FP (2000) Early detection of developmental and behavioral problems. Pediatrics in Review 21: 272-280.

34. Eyberg S (1980) The Eyberg Child Behavior Inventory. Journal of Clinical Child Psychology 9: 22-28.

35. Mouton-Simien P, McCain AP, Kelley ML (1997) The development of the toddler behavior screening inventory. Journal of Abnormal Child Psychology 25: 59-64.

36. Bornstein MH, Hahn CS, Haynes OM (2010) Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: Developmental cascades. Development and Psychopathology 22: 717-735.

37. Yirmiya N, Charman T (2010) The prodrome of autism: early behavioral and biological signs, regression, peri- and post-natal development and genetics. Journal of Child Psychology and Psychiatry 51: 432-458.

38. Kousemaker NPJ (1996) Zoeken, vinden, zorgen delen: de ontwikkeling van een praktijkparadigma voor onderkenning en pedagogische preventie van psychosociale problematiek in de Jeugdgezondheidszorg [Searching, finding, sharing care; the development of a practise paradigm for the descernment and pedagogical prevention of psychosocial problems in Youth Healthcare].

39. Hermans J, Ory F, Schrijvers G (Julius Centrum: Utrecht; 2005) Supporting development and parenting: sooner, faster and better: an advice for early detection and interventions in regard to developmental and parenting problems.

40. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.

41. Briggs-Gowan MJ, Carter AS (2002) Brief Infant-Toddler Social and Emotional Assessment (BITSEA) mannual, version 2.0. New Haven: CT:Yale University.

42. Haapsamo H, Ebeling H, Soini H, Joskitt L, Larinen K, et al. (2009) Screening infants with social and emotional problems: A pilot study of the Brief Infant Toddler Social and Emotional Assessment (BITSEA) in Northern Finland. International Journal of Circumpolar Health 68: 386-393.

43. Karabekiroglu K, Briggs-Gowan MJ, Carter AS, Rodopman-Arman A, Akbas S (2010) The clinical validity and reliability of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). Infant Behavior & Development 33: 503-509.

44. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, et al. (2009) The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behavior & Development 32: 291-297.

45. Wendland J, Danet M, Gacoin E, Didane N, Bodeau N, et al. (2014) French Version of the Brief Infant-Toddler Social and Emotional Assessment Questionnaire-BITSEA. J Pediatr Psychol.

PART ONE

The psychometric properties of early detection instruments

Chapter 2

Reliability and validity of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA)

Ingrid Kruizinga
Wilma Jansen
Carolien L. de Haan
Jan van der Ende
Alice S. Carter
Hein Raat

Published as:
Kruizinga I, Jansen W, de Haan CL, van der Ende J, Carter AS, et al. (2012) Reliability and Validity of the Dutch Version of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). PLoS ONE 7(6): e38762. doi:10.1371/journal.pone.0038762

ABSTRACT

Background: The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) is a relatively new and short (42-item) questionnaire that measures psychosocial problems in toddlers and consists of a Problem and a Competence scale. In this study the reliability and validity were examined for the whole group and for gender and ethnicity subgroups.

Methods: Parents of 7140 two-year-old children were invited in the study, of which 3170 (44.4%) parents completed the BITSEA. For evaluation of the score distribution, the presence of floor/ceiling effects was determined. The internal consistency (Cronbach's alpha) was evaluated and in subsamples the test-retest, parent-childcare provider interrater reliability and concurrent validity with regard to the Child Behavioral Checklist (CBCL). Discriminative validity was evaluated by comparing scores of parents that worry and parents that do not worry about their child's development.

Results: The BITSEA showed no floor or ceiling effects. Psychometric properties of the BITSEA Problem and Competence scale were respectively: Cronbach's alphas were 0.76 and 0.63. Test-retest correlations were 0.75 and 0.61. Interrater reliability correlations were 0.30 and 0.17. Concurrent validity was as hypothesized. The BITSEA was able to discriminate between parents that worry about their child and parents that do not worry. The psychometric properties of the BITSEA were comparable across gender and ethnic background.

Conclusion: The results in this large-scale study of a diverse sample support the reliability and validity of the BITSEA Problem scale. The BITSEA Competence scale needs further study. The performance of the BITSEA appears to be similar in subgroups by gender and ethnic background.

INTRODUCTION

Psychosocial problems, such as social-emotional and behavioural problems, are prevalent among 12% to 16% of two-year-old children [1]. Psychosocial problems in preschool aged children are associated with disorders later in life, such as oppositional defiant disorder, attention deficit disorder, conduct disorder, simple phobia, avoidant disorder and depressive disorder NOS [2,3]. Measurement, early detection and treatment of psychosocial problems at a young age is important because this may contribute to a reduction of problems and an increase of competencies at older ages [4,5]. To measure psychosocial problems, reliable and valid instruments are necessary.

Short comprehensive instruments that are appropriate to measure psychosocial problems in children of preschool age are limited [6]. Existing instruments, such as the Eyberg Child Behavior Inventory [7] or the Toddler Behavior Screening Inventory [8], only measure problem behaviour and do not address social-emotional competencies. Measuring delays in social-emotional competence, however, is also important since delays in competence are for instance related to internalising and externalising problems later in life [9]. There remains a need for a short instrument that measures both problems and delays in competence.

The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) [10], developed in the United States of America, is a short (42-item) questionnaire measuring psychosocial problems and delays in the acquisition of competence in toddlers. The BITSEA consists of a Problem scale and a Competence scale, and can be used in epidemiological studies, in (preventive) child health care and in the early intervention settings for children between the ages of 12 and 36 months [10,11]. The BITSEA is a shorter version of the Infant-Toddler Social and Emotional Assessment (ITSEA) [12,13], which has been reported to have an acceptable factor structure, test-retest reliability, interrater reliability and validity in (a) community samples [13], (b) a sample of young children referred to an early intervention program [11] and (c) a clinical sample of young children referred for psychiatric assessment [14].

Only a few studies have evaluated the reliability and validity of the BITSEA [10,15,16]. The objective of this study was to investigate the following psychometric properties of the Dutch version of the BITSEA in a large sample of preschool children in the Netherlands:

1. the score distribution of the BITSEA;
2. the reliability of the BITSEA scales scores (internal consistency, test-retest reliability and interrater reliability);

3. the validity of the BITSEA scales interpretation (concurrent validity and discriminative validity).

Additionally we evaluated the score distribution, reliability and validity within subgroups of boys and girls, as well as native and immigrant children, because psychometric properties might differ between these subgroups [17-19].

METHODS

Ethics statement

Part of the data became available in the context of the government approved routine health examinations of the preventive child health care. Separate informed consent was therefore not requested. Only anonymous data were used and the questionnaires were completed on a voluntary basis. Parents received written information on these questionnaires and were free to object to participation. Observational research with data does not fall within the ambit of the Dutch Act on research involving human subjects and does not require the approval of an ethics review board. As part of the data was anonymous for the researchers, this part of the study is not covered by the WMA Declaration of Helsinki. Informed consent was obtained for participation for the test-retest and interrater reliability data-collection, since these data were not anonymous and not part of the routine health examinations. This part of the study has been conducted according to the principles expressed in the WMA Declaration of Helsinki. The Medical Ethics Committee of the Erasmus Medical Centre Rotterdam approved the study protocol and consent procedures.

Data collection

The present study was embedded in broader examinations of the BITSEA as an early detection tool of psychosocial problems in toddlers and has been described in detail elsewhere [20]. The present study was conducted in the larger Rotterdam area in the Netherlands among two-year-old children and their parents, who were invited between April 2010 and April 2011 by child health care organizations for well-child visits: A few weeks before the well-child visit was scheduled, parents of 7140 children received a child health monitor questionnaire by mail, including among others the BITSEA and Child Behavioral Checklist (CBCL1.5-5) and written information about the study. Parents decided for themselves whether the father

or mother would complete the questionnaire. The parent-completed BITSEA was used by a child health professional during the well-child visit to assess the development of the child. Parents of 3320 (46.5%) children attended the well-child visit; 53.5% of invited parents did not attend the well-child visit and did not complete the questionnaire. Of those parents that did attend the well-child visit, 3170 (95.5%) handed in the completed child health monitor questionnaire. Children were excluded from the analyses if there were too many missing items (Problem scale >5, Competence scale >2) on both BITSEA scales (n=43) [21], leaving a study population of 3127 (94.2%) children. The CBCL1.5-5 [22] was also included in the child health monitor questionnaire but only for research purposes (i.e. evaluating the concurrent validity of the BITSEA). Parents of 2304 (69.4%) children wanted to contribute to the study and also completed the CBCL1.5-5.

Test-retest and interrater reliability was evaluated in the subsample of parents that completed the child health monitor questionnaire in the month prior to receiving the questionnaire by the researchers. A subgroup of 314 parents were mailed the BITSEA again to assess the test-retest reliability which resulted in a response by parents of 120 (38.2%) children. The range of the period between completion of questionnaires was 13-77 days (mean=44.7, SD=18.1). Additionally, BITSEA questionnaires were mailed to childcare providers (i.e. child day care facilities outside home) of a subgroup of 130 children to assess interrater reliability, which resulted in a response of 75 (57.7%) completed questionnaires. The range of the period between completion of questionnaires was 3-76 days (mean=45.8, SD=21.5).

Measures

The BITSEA consists of 42 items with three response options ('not true/rarely', 'somewhat true/sometimes', 'very true/often'). Versions are available for parents and childcare providers. The childcare provider form is almost identical to the parent form but has some wording adaptations to make it appropriate for the childcare setting. The BITSEA is comprised of two multi-item scales, a Problem scale (31 items) and Competence scale (11 items), and responses can be summed for each scale. The possible score range of the Problem scale is 0-62 and of the Competence scale 0-22. A high score on the Problem scale or a low score on the Competence scale is less favourable [21]. In addition to the 42 items, the BITSEA has two single-item questions on parent worries regarding child language development and child behaviour, emotions or relationships. The BITSEA was translated into Dutch according to international guidelines [23].

In addition to the BITSEA, the CBCL1.5-5 was completed by parents in order to evaluate the concurrent validity of the BITSEA. The well-validated [22] 100-item CBCL1.5-5 is designed for children aged 18 months to 5 years and has two domains (Internalising and Externalising) and a Total Problem score. Answers are given on a 3-point scale ('not true', 'somewhat or sometimes true' and 'very true or often true').

Items on standard socio-demographic variables were included; which parent completed the questionnaire, ages of parents and child, child gender, child and parents' country of birth, parents' educational level and employment status, and family composition. A child was considered native if both parents were born in the Netherlands, a child was considered an immigrant if at least one of the parents is born outside the Netherlands [24].

Analyses

Analyses were performed with SPSS 19.0 (SPSS Inc. 2010). Differences in mean BITSEA scores between boys and girls and between native and immigrant children were tested with independent sample t-tests.

Score distribution

Score distribution was evaluated by assessing the presence of floor and ceiling effects (i.e. >15% of the respondents have the minimal and/or maximal score) [25], mean scale scores and the 25th, 50th and 75th percentile points.

Reliability

Cronbach's alpha was used to evaluate the internal consistency of the Problem and Competence scales. An alpha of 0.70 or higher is considered acceptable [26]. Differences in internal consistency across gender and ethnic background subgroups was tested by computing critical F-statistics [27] with alpha set to 0.01. Test-retest and interrater reliability of the BITSEA-scales were assessed with the Intraclass Correlation Coefficients (ICC), using a two-way random effect model with absolute agreement. An ICC of 0.70 or higher is considered to indicate acceptable test-retest and interrater reliability [25]. To test the difference between gender and ethnic background subgroups for test-retest and interrater reliability, ICC Fisher r-to-z transformations were performed and a two-tailed criterion for significance was used.

Validity

Concurrent validity was evaluated by assessing Pearson correlations between BITSEA and CBCL1.5-5 scale scores. Concurrent validity is hypothesized to be expressed in large positive correlations and small to medium negative correlations between respectively BITSEA Problem and Competence scales with the CBCL1.5-5 Internalising, Externalising and Total Problem scores. A correlation of 0.1 is considered small, 0.3 is considered medium and >0.5 is considered large [28].

Discriminative validity is evaluated by assessing the ability of the BITSEA to discriminate between a subgroup *without* parents who reported worries about their child's behaviour, emotions or relationships and a subgroup *with* parents who reported worries about their child's behaviour, emotions or relationships. This single-item question is part of the BITSEA, however does not add to either BITSEA scale score, therefore we regarded this question as suitable to evaluate discriminative validity. We hypothesized that discriminative validity will be reflected in less favourable BITSEA scores for children of parents with worries about their child [29]. Differences in mean BITSEA scores between these groups were tested with an independent sample t-test and effect sizes were defined as $d = | [mean_{not\ worried} - mean_{worried}] / SD_{worried} |$ [30]; $0.20 \leq d < 0.50$ indicates a small effect, $0.50 \leq d < 0.80$ indicates a medium effect and $d \geq 0.80$ indicates a large effect. Discriminative validity, as described above, was also evaluated by gender and ethnic background subgroups. We hypothesized that we would find the same pattern of results within subgroups as in the general population.

RESULTS

Mean child age was 23.7 months (SD=0.7), 48.9% were girls, and 55.7% of the children had a Dutch ethnic background. Mean age of the mother was 33.5 years (SD=5.1) and mean age of the father was 36.3 years (SD=5.5). In 88.1% of the cases the mother or both parents were the respondent(s). See Table 2.1 for more information on demographic characteristics of the study population.

Score distribution

Floor and ceiling effects were absent (Table 2.2). Mean scale scores and the 25th, 50th and 75th percentile points are presented in Table 2.2.

Table 2.1 Characteristics of the study population, N=3127

	% of participants	Mean (SD)
Mother characteristics		
Age (years)		33.5 (5.1)
Country of birth (The Netherlands)	65.1	
Educational level[1]		
Lower general education or less	23.4	
Intermediate vocational/pre-university	30.5	
Higher vocational /university	39.6	
Employment[1]		
Employed	63.7	
Homemaker	16.5	
Unemployed	9.9	
Father characteristics		
Age (years)		36.3 (5.5)
Country of birth (The Netherlands)	61.4	
Educational level[1]		
Lower general secondary or less	21.6	
Intermediate vocational/pre-university	27.7	
Higher vocational /university	36.2	
Employment[1]		
Employed	79.3	
Homemaker	0.8	
Unemployed	6.6	
Child characteristics		
Age (months)		23.7 (0.7)
Gender (girls)	48.9	
Ethnic background[2] (native)	55.7	
Family characteristics		
Two-parent household	82.5	
One-child family	42.1	
Respondent (mother or both parents)	88.1	

[1] Percentages do not sum to 100 because of missing values.
[2] A child is considered native when both parents were born in The Netherlands.

Boys had both a significantly higher mean Problem score (8.2, SD=5.6) compared to girls (7.4, SD=4.9), $p<0.01$, and a significantly lower mean Competence score (17.1, SD=3.0) compared to girls (17.9, SD=3.0), $p<0.01$. Immigrant children had both a significantly higher mean Problem score (9.3, SD=5.9) compared to native children (6.7, SD=4.4), $p<0.01$, and a significantly lower mean Competence score (16.7, SD=3.3) compared to native children (18.1, SD=2.7), $p<0.01$ (Table 2.2).

Table 2.2 Score distributions and internal consistency of BITSEA-scales, as reported by the parents, by gender and ethnic background, N=2237

BITSEA scales	Mean score (SD)	Range	% min[1]	% max[1]	25th %tile	50th %tile	75th %tile	Cronbach's alpha[2]
Total N=2237								
Problem	7.8 (5.3)	0-40	1.8	0.0	4	7	10	0.76
Competence	17.5 (3.0)	0-22	0.1	5.6	16	18	20	0.63
Boys N=1124								
Problem	8.2[a] (5.6)	0-40	1.6	0.0	4	7	11	0.77
Competence	17.1[a] (3.0)	1-22	0.0	4.0	15	17	19	0.61
Girls N=1098								
Problem	7.4[a] (4.9)	0-30	2.1	0.0	4	6	10	0.74
Competence	17.9[a] (3.0)	0-22	0.1	7.4	16	18	20	0.65
Native children N=1354								
Problem	6.7[b] (4.4)	0-40	2.1	0.0	3	6	9	0.70
Competence	18.1[b] (2.7)	4-22	0.0	7.6	16	19	20	0.60
Immigrant children N=883								
Problem	9.3[b] (5.9)	0-38	1.5	0.0	5	8	12	0.78
Competence	16.7[b] (3.3)	0-22	0.1	3.1	15	17	19	0.64

[1] % of respondents with the lowest (min) and highest (max) BITSEA scale score (ceiling/floor).
[2] No significant differences between subgroups by gender or ethnic background in internal consistency, p>0.01.
a = significant difference in mean BITSEA scores between boys and girls, p<0.01.
b = significant difference in mean BITSEA scores between native and immigrant children, p<0.01.

Reliability

Internal consistency was 0.76 for the Problem scale and 0.63 for the Competence scale (Table 2.2). Test-retest reliability was 0.75 for the Problem scale and 0.61 for the Competence scale (Table 2.3). Parent/childcare provider interrater reliability was 0.30 for the Problem scale and 0.17 for the Competence scale (Table 2.3). No significant differences in reliability indices for gender and ethnic background subgroups were found.

Validity

Concurrent validity: The BITSEA Problem scale was positively correlated with the CBCL1.5-5, Pearson coefficients of 0.66 (Internalising), 0.65 (Externalising) and 0.75 (Total Problem). The BITSEA Competence scale was negatively correlated with the CBCL1.5-5, Pearson coefficients of -.26 (Internalising), -0.23 (Externalising) and -0.26 (Total Problem). All

Table 2.3 Test-retest and parent/childcare provider interrater reliability for BITSEA scores, by gender and ethnic background

BITSEA scales		Problem			Competence		
		Test-retest reliability					
Sample	N	ICC	Mean (SD) test	Mean (SD) retest	ICC	Mean (SD) test	Mean (SD) retest
Total	120	0.75	7.1 (4.6)	6.6 (4.4)	0.61	18.1 (2.8)	18.1 (3.0)
Boys	63	0.70	7.5 (4.9)	7.0 (4.6)	0.66	17.4 (3.1)	17.3 (3.4)
Girls	57	0.82	6.6 (4.2)	6.2 (4.2)	0.41	18.8 (2.2)	19.1 (2.2)
Native children	87	0.66	6.4 (3.7)	6.2 (3.9)	0.57	18.3 (2.4)	18.5 (2.6)
Immigrant children	33	0.83	8.9 (6.0)	7.9 (5.4)	0.65	17.5 (3.5)	17.0 (3.6)
		Interrater reliability					
Sample	N	ICC	Mean (SD) parent	Mean (SD) childcare provider	ICC	Mean (SD) parent	Mean (SD) childcare provider
Total	75	0.30	7.3 (4.5)	7.5 (3.9)	0.17	18.6 (2.6)	16.1 (3.7)
Boys	38	0.34	7.1 (4.9)	7.4 (3.3)	0.19*	18.6 (2.3)	16.2 (3.3)
Girls	37	0.26*	7.5 (4.0)	7.6 (4.6)	0.16*	18.6 (2.9)	16.1 (4.2)
Native children	56	0.33	7.2 (4.6)	7.2 (4.1)	0.16*	18.8 (2.3)	16.4 (3.6)
Immigrant children	19	0.17*	7.5 (4.2)	8.3 (3.2)	0.15*	17.8 (3.3)	15.4 (4.2)

Note: There are no significant differences in ICC between gender and ethnic background.
* ICC is not significant, $p>0.05$.

correlations were significant, $p<0.01$. A similar pattern of correlations between BITSEA and CBCL1.5-5 was found for gender and ethnic background subgroups (Table 2.4).

Discriminative validity: BITSEA scores of 482 (15.2%) children of parents who were worried were compared to BITSEA scores of 2621 (82.7%) children of parents that were not worried (percentages do not sum to 100% because of missing values). The mean BITSEA Problem score was higher in the 'worried subgroup' compared to the 'not worried subgroup', respectively mean=12.8 (SD=6.3) and mean=6.9 (SD=4.5), $p<0.01$, effect size=0.93. BITSEA Competence scores were lower in the 'worried subgroup' compared to the 'not worried subgroup', respectively mean=16.0 (SD=3.5) and mean=17.8 (SD=2.8), $p<0.01$, effect size=0.52. A similar pattern of differences in mean BITSEA scores between 'worried' parents and 'not worried' parents was found for gender and ethnic background subgroups (Table 2.5).

Table 2.4 Concurrent validity (BITSEA and CBCL1.5-5) by gender and ethnic background, N=2304

Pearson correlation	Total		Boys		Girls		Native children		Immigrant children	
	Problem	Competence	Problem	Competence	Problem	Competence	Problem	Competence	Problem	Competence
CBCL scales										
Internalising	0.66	-0.26	0.65	-0.28	0.66	-0.24	0.60	-0.23	0.66	-0.20
Externalising	0.65	-0.23	0.67	-0.24	0.64	-0.21	0.64	-0.28	0.66	-0.13
Total Problems	0.75	-0.26	0.75	-0.29	0.75	-0.34	0.72	-0.27	0.75	-0.17

Note: All correlations are significant, $p<0.01$.

Table 2.5 Discriminative ability of the BITSEA between subgroups differing in parental worries about a child's behaviour, emotions or relationships, by gender and ethnic background, N=3103

Mean (SD)	Total			Boys			Girls			Native children			Immigrant children		
	Parental worries			Parental worries			Parental worries			Parental worries			Parental worries		
BITSEA scales	Not worried N=2621	Worried N=482	Effect size d[1]	Not worried N=1301	Worried N=253	Effect size d[1]	Not worried N=1300	Worried N=224	Effect size d[1]	Not worried N=1574	Worried N=173	Effect size d[1]	Not worried N=1047	Worried N=309	Effect size d[1]
Problem	6.9 (4.5)	12.8 (6.3)	0.93[a]	7.3 (4.7)	13.3 (6.9)	0.87[a]	6.6 (4.3)	12.2 (5.5)	1.02[a]	6.2 (3.9)	11.4 (5.5)	0.95[a]	8.1 (5.1)	13.6 (6.6)	0.84[a]
Competence	17.8 (2.8)	16.0 (3.5)	0.52[b]	17.4 (2.9)	15.6 (3.4)	0.54[b]	18.2 (2.7)	16.5 (3.6)	0.48[c]	18.3 (2.6)	16.9 (3.1)	0.46[c]	17.1 (3.0)	15.5 (3.7)	0.44[c]

Note: All differences in means between 'worried' and 'not worried' are significant at $p<0.01$ for the total sample and subgroups.

[1] Difference of the means divided by SD in the subgroup with worries.

[a] indicates a large effect ($d\geq0.8$).

[b] indicates a medium effect ($0.5\leq d\geq0.8$).

[c] indicates a small effect ($0.2\leq d\geq0.5$).

DISCUSSION

The present study evaluated the psychometric properties of the Dutch version of the BITSEA in a large community sample in the Netherlands with a focus on differences across child gender and child ethnic background subgroups. The following psychometric properties of the BITSEA were determined in the present study: internal consistency, interrater reliability, test-retest reliability, concurrent validity and discriminative validity. The BITSEA Problem scale showed acceptable performance on all psychometric properties, whereas the BITSEA Competence scale showed acceptable performance on concurrent and discriminative validity. There were no differences in the psychometric properties of the BITSEA between boys and girls or between native and immigrant children.

Score distribution

The BITSEA showed no floor or ceiling effects, which means that changes within toddlers with very low or very high scores can be measured. It also means that a toddler with a low score can be differentiated from other toddlers with low scores and that a toddler with a high score can be differentiated from other toddlers with high scores [25].

Reliability

Internal consistency for the Problem scale was adequate (>0.70), but the internal consistency for the Competence scale was marginal (i.e. 0.63). Lower internal consistency for the Competence scale might be explained by inclusion of some items that assess behaviours that may not be expected to co-occur in young children, and items that are likely to show limited variability because they address early emerging competencies to identify significant social competence delays [10].

Test-retest reliability was adequate (>0.70) for the Problem scale and marginal (i.e. 0.61) for the Competence scale. These results mean that the BITSEA Problem scale provides stable outcomes over time, assuming that no real changes in psychosocial problems occur.

Interrater reliability was lower than the suggested guideline of 0.70. However, an interrater reliability meta-analysis of 119 studies, in which 26 studies reported interrater reliability between parent and teacher, found a mean correlation of 0.27 [31]. Correlations between parents and childcare provider/teacher are typically lower than correlations between parents.

Lower correlations between measures of different observers can partly be explained by different settings in which a child is observed [31]. Compared to the mean reported parent-teacher interrater reliability, the Problem scale interrater reliability in this study was typical. However, the interrater reliability of the Competence scale was much lower than 0.27 and raises concerns about the reliability of this measure.

Validity

As hypothesized, the BITSEA showed good concurrent validity; the BITSEA Problem scale had a strong positive correlation with CBCL1.5-5 Internalising, Externalising and Total Problem scores. Also as hypothesized, the BITSEA Competence scale had a negative correlation of medium strength with CBCL1.5-5 Internalising, Externalising and Total Problem scores.

The BITSEA scores were able to distinguish between parents reporting worry about their child's behaviour, emotions or relationships and parents who were not worried, indicating a good discriminative validity. Previous research illustrated a strong relationship between parents' concerns and children's developmental status [29], which supports our findings on the discriminative reliability of the BITSEA.

Mean BITSEA scores were less favourable for boys compared to girls, and for immigrant children compared to native children. These findings are in line with previous studies that report boys experience psychosocial problems more often than girls [32] and that psychosocial problems are more often reported in immigrant children compared to native children [33,34].

The psychometric properties in this study are largely in line with what was found in previous studies on the BITSEA [10,16]. One study found slightly higher internal consistency [16], another study found higher interrater reliability on the Competence scale and test-retest reliability [10] compared to our results. Differences in psychometric properties of the BITSEA may be explained by different social demographic characteristics and a different setting (e.g. in the other studies the BITSEA was not used by a child health professional to assess the child's development).

Our study has a few limitations. First, is that in the current study we have no data on the large non-response group. No information is available on parents that did not attend the well-child visit. It might be possible that parents avoid attending the well-child visit because they are afraid of possible interventions from Youth Care, but it might also be possible that

parents do not find it necessary to attend the well-child visit because they feel confident that their child has no problems. Because the characteristics of the parents that are missed are unknown, it is unclear how the non-response has influenced the results on the psychometric properties of the BITSEA. However, we found no differences in psychometric properties within subgroups, so therefore we are confident that the non-response did not have a large impact on the outcomes. Second, the report by parents introduces the proxy-problem; self-report by two-year-old children on their psychosocial problems is not possible, because children of this age lack the necessary language skills and the cognitive abilities to interpret the questions and they do not have a long-term view of events [35]. Therefore, proxy by parents may be a useful alternative [36].

A major strength of our study is the large and diverse sample size. Additionally, the setting in which the respondents were invited to complete the BITSEA, the daily practice of well-child visit at the child health care centre, can be seen as either a strength or a limitation. We evaluated the psychometric properties in a setting in which the BITSEA might be implemented; however this specific setting might, on the other hand, hamper generalisations of our results to other settings.

We recommend future studies to evaluate the psychometric properties of the BITSEA in a different sample and setting. The setting in this study was the daily practice of a well-child visit in an urban area; but it would be good to be able to replicate these results in a more rural area, possibly outside the context of a well-child visit. Also, we recommend future studies to evaluate the BITSEA as an early detection tool for psychosocial problems in toddlers (i.e. the ability of the BITSEA to correctly classify children with and without psychosocial problems) for which the sensitivity and specificity of the BITSEA should be evaluated using a clinical sample of children with a diagnosis made by a professional [37]. Furthermore, referrals by child health professionals based on BITSEA scores and subsequent use of the (mental) health care system of children should also be investigated.

In conclusion, the results of our study support the reliability and validity of the BITSEA Problem scale. Further studies regarding the reliability of the Competence scale are advised. The performance of the BITSEA appears to be similar in boys and girls and in native and immigrant children. The BITSEA is a promising instrument to measure psychosocial problems in toddlers.

ACKNOWLEDGMENTS

We would like to express our gratitude to Irene Hoekstra and Mirjana Hillebrand for their assistance in this study.

REFERENCES

1. Briggs-Gowan MJ, Carter AS, Skuban EM, Horwitz SM (2001) Prevalence of social-emotional and behavioral problems in a community sample of 1-and 2-year-old children. Journal of the American Academy of Child and Adolescent Psychiatry 40: 811-819.

2. Mesman J, Koot HM (2001) Early preschool predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry 40: 1029-1036.

3. Lavigne JV, Arend R, Rosenbaum D, Binns HJ, Christoffel KK, et al. (1998) Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry 37: 1246-1254.

4. Durlak JA, Wells AM (1998) Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. American Journal of Community Psychology 26: 775-802.

5. Elliot J, Prior M, Merrigan C, Ballinger K (2002) Evaluation of a community intervention programme for preschool behaviour problems. Journal of Paediatrics and Child Health 38: 41-50.

6. Glascoe FP (2000) Early detection of developmental and behavioral problems. Pediatrics in Review 21: 272-280.

7. Eyberg S (1980) The Eyberg Child Behavior Inventory. Journal of Clinical Child Psychology 9: 22-28.

8. Mouton-Simien P, McCain AP, Kelley ML (1997) The development of the toddler behavior screening inventory. Journal of Abnormal Child Psychology 25: 59-64.

9. Bornstein MH, Hahn C-S, Haynes OM (2010) Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: Developmental cascades. Development and Psychopathology 22: 717-735.

10. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.

11. Briggs-Gowan MJ, Carter AS (2007) Applying the infant-toddler social & emotional assessment (ITSEA) and brief-itsea in early intervention. Infant Mental Health Journal 28: 564-583.

12. Briggs-Gowan MJ, Carter AS (1998) Preliminary acceptability and psychometrics of the Infant-Toddler Social and Emotional Assessment (ITSEA): A new adult-report questionnaire. Infant Mental Health Journal 19: 422-445.

13. Carter AS, Briggs-Gowan MJ, Jones SM, Little TD (2003) The Infant-Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. Journal of Abnormal Child Psychology 31: 495-514.

14. Visser JC, Smeekens S, Rommelse N, Verkes RJ, Van der Gaag RJ, et al. (2010) Assessment of psychopathology in 2-to5-year-olds: Applying the Infant-Toddler Social Emotional Assessment. Infant Mental Health Journal 31: 611-629.

15. Haapsamo H, Ebeling H, Soini H, Joskitt L, Larinen K, et al. (2009) Screening infants with social and emotional problems: A pilot study of the Brief Infant Toddler Social and Emotional Assessment (BITSEA) in Northern Finland. International Journal of Circumpolar Health 68: 386-393.

16. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, et al. (2009) The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behavior & Development 32: 291-297.

17. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, et al. (2002) Assessing health status and quality-of-life instruments: Attributes and review criteria. Quality of Life Research 11: 193-205.

18. Ohan JL, Johnston C (2005) Gender appropriateness of symptom criteria for attention-deficit/hyperactivity disorder, oppositional-defiant disorder, and conduct disorder. Child Psychiatry & Human Development 35: 359-381.

19. Warnecke RB, Johnson TP, Chavez N, Sudman S, Orourke DP, et al. (1997) Improving question wording in surveys of culturally diverse populations. Annals of Epidemiology 7: 334-342.

20. Kruizinga I, Jansen W, Carter AS, Raat H (2011) Evaluation of an early detection tool for socialemotional and behavioral problems in toddlers: The Brief Infant Toddler Social and Emotional Assessment - A cluster randomized trial. BMC Public Health 11:494.

21. Briggs-Gowan MJ, Carter AS (2002) Brief Infant-Toddler Social and Emotional Assessment (BITSEA) mannual, version 2.0. New Haven: CT:Yale University.

22. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington:VT: University ofVermont, Research Center for Children, Youth, and Families.

23. Wild D, Grove A, M Martin, al. e (2005) Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value in Health 8: 94-104.

24. Definitions 'native' and 'immigrant' (2012) - http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?conceptid=37. [Statistics Netherlands].

25. Terwee CB, Bot SDM, de Boer MR, van der Windt D, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 60: 34-42.

26. Nunnally JC, Bernstein IH (1994) Psychometric theory. New York: McGraw-Hill.

27. Glanz S (1992) Primer of Biostatistics. New York: McGraw-Hill.

28. Cohen J (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erblaum Associates.

29. Glascoe FP (1997) Parents'concerns about children's development: Prescreening technique or screening test? Pediatrics 99: 522-528.

30. Cohen J (1977) Statistical power analysis for the behavioral sciences. New York: Academic Press.

31. Achenbach TM, McConaughy SH, Howell CT (1987) Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin 101: 213-232.

32. Kraemer S (2000) The fragile male. British Medical Journal 321: 1609-1612.

33. Reijneveld S, Harland P, Brugman E, Verhulst F, Verloove-Vanhorick S (2005) Psychosocial problems among immigrant and non-immigrant children - Ethnicity plays a role in their occurrence and identification. European Child & Adolescent Psychiatry 14: 145-152.

34. Vollebergh WAM, ten Have M, Dekovic M, Oosterwegel A, Pels T, et al. (2005) Mental health in immigrant children in the Netherlands. Social Psychiatry and Psychiatric Epidemiology 40: 489-496.

35. Vogels T, Verrips GHW, Verloove-Vanhorick SP, Fekkes M, Kamphuis RP, et al. (1998) Measuring health-related quality of life in children: the development of the TACQOL parent form. Quality of Life Research 7: 457-465.

36. Theunissen NCM, Vogels TGC, Koopman HM, Verrips GHW, Zwinderman KAH, et al. (1998) The proxy problem: child report versus parent report in health-related quality of life research. Quality of Life Research 7: 387-397.

37. Glascoe FP (2005) Screening for developmental and behavioral problems. Mental Retardation and Developmental Disabilities Research Reviews 11: 173-179.

Chapter 3

Reliability and validity of the KIPPPI: an early detection tool for psychosocial problems in toddlers

Ingrid Kruizinga
Wilma Jansen
Carolien L. de Haan
Hein Raat

Published as:
Kruizinga I, Jansen W, de Haan CL, Raat H (2012) Reliability and Validity of the KIPPPI: An Early Detection Tool for Psychosocial Problems in Toddlers. PLoS ONE 7(11): e49633. doi:10.1371/journal.pone.0049633

ABSTRACT

Background: The KIPPPI (Brief Instrument Psychological and Pedagogical Problem Inventory) is a Dutch questionnaire that measures psychosocial and pedagogical problems in 2-year olds and consists of a KIPPPI Total score, Wellbeing scale, Competence scale, and Autonomy scale. This study examined the reliability, validity, screening accuracy and clinical application of the KIPPPI.

Methods: Parents of 5959 2-year-old children in the Rotterdam area, the Netherlands, were invited to participate in the study. Parents of 3164 children (53.1% of all invited parents) completed the questionnaire. The internal consistency was evaluated and in subsamples the test-retest reliability and concurrent validity with regard to the Child Behavioral Checklist (CBCL). Discriminative validity was evaluated by comparing scores of parents who worried about their child's upbringing and parent's that did not. Screening accuracy of the KIPPPI was evaluated against the CBCL by calculating the Receiver Operating Characteristic (ROC) curves. The clinical application was evaluated by the relation between KIPPPI scores and the clinical decision made by the child health professionals.

Results: Psychometric properties of the KIPPPI Total score, Wellbeing scale, Competence scale and Autonomy scale were respectively: Cronbach's alphas: 0.88, 0.86, 0.83, 0.58. Test-retest correlations: 0.80, 0.76, 0.73, 0.60. Concurrent validity was as hypothesized. The KIPPPI was able to discriminate between parents that worried about their child and parents that did not. Screening accuracy was high (>0.90) for the KIPPPI Total score and for the Wellbeing scale. The KIPPPI scale scores and clinical decision of the child health professional were related ($p<0.05$), indicating a good clinical application.

Conclusion: The results in this large-scale study of a diverse general population sample support the reliability, validity and clinical application of the KIPPPI Total score, Wellbeing scale and Competence scale. Also, the screening accuracy of the KIPPPI Total score and Wellbeing scale were supported. The Autonomy scale needs further study.

INTRODUCTION

The importance of early detection of psychosocial problems, such as social-emotional and behavioural problems in toddlers is increasingly recognized [1-3]. In the Netherlands, in approximately 8-9 percent of preschool children, child health professionals identify psychosocial problems [4,5]. Psychosocial problems are associated with psychological disorders later in life [6,7]. Therefore, it is important to detect and treat psychosocial problems at a young age, because early detection and treatment may contribute to a reduction of problems and an increase in competencies at an older age [8,9]. However, studies show that a relatively small number of children with psychosocial problems are identified by child health professionals (i.e. 29% of the children who scored in the clinical range of the CBCL Total Problem score) [5] and are being referred to mental health services (i.e. 13% of the children who scored in the clinical range of the CBCL Total Problem score) [10]. The accuracy of identification of psychosocial problems should be enhanced [11]. To facilitate early detection of psychosocial problems in toddlers, child health professionals can use reliable and valid parent-completed questionnaires [12-14].

The toddler KIPPPI [15] (KIPPPI is a Dutch acronym for Brief Instrument Psychological and Pedagogical Problem Inventory) was developed in the Netherlands and measures psychosocial problems in 2-year olds, which might be possible pedagogical challenges for the parents. The KIPPPI has 67 items and consist of a Wellbeing scale (31 items), Competence scale (25 items) and an Autonomy scale (11 items). The KIPPPI Total score is the sum of the scale scores. Child health professionals use the KIPPPI as an early detection tool during well-child visits to assess the child's psychosocial problems that might also be related to pedagogical problems. The KIPPPI is specifically developed for use in the preventive child health care and is widely used in the Netherlands. As many aspects of psychosocial problems are addressed in the questionnaire, the KIPPPI can be used by the child health professional to guide conversation with the parent.

Little is known about the reliability and validity of the KIPPPI. The objective of this study was to determine in a large general population sample of 2-year old children: the score distribution of the KIPPPI (mean KIPPPI scores and standard deviations for the total population as well as for subgroups by child's gender and ethnicity; and floor and ceiling effects) and the following psychometric properties of the KIPPPI; the reliability of the KIPPPI scale scores (internal consistency and test-retest reliability); the validity of the KIPPPI scale interpretation (concurrent validity and discriminative validity); the screening accuracy of the KIPPPI was

evaluated relative to the Child Behavioral Checklist 1.5-5 (CBCL1.5-5), a well-validated questionnaire that measures behavioural, emotional and social problems in preschool children. Additionally we evaluated the clinical application; whether the KIPPPI scores were related to the clinical decision of the child health professionals.

METHODS

Ethics statement

Part of the data became available in the context of the government-approved routine health examinations of the preventive child health care. Anonymous data were used in this study and the questionnaires were completed on a voluntary basis. Parents received written information on the study and were free to object to participation. Observational research does not fall within the ambit of the Dutch Act on research involving human subjects and does not require the approval of an ethics review board. Written informed consent was obtained from a subgroup of parents that participated in a substudy to evaluate the test-retest reliability and to compare KIPPPI scores with CBCL1.5-5 scores, because these data were not anonymous and were not collected as part of the routine health examinations. The study was conducted in accordance with the WMA Declaration of Helsinki principles. We received a formal waiver (i.e. declaration of no objection) from The Medical Ethics Committee of the Erasmus Medical Center Rotterdam.

Data collection

The present study was conducted among two-year-old children and their parents, who were invited between April 2010 and April 2011 by child health care organizations in the larger Rotterdam area, the Netherlands, for well-child visits: A few weeks before the well-child visit was scheduled, parents of 5959 children were sent a child health monitor questionnaire by mail, including among others the KIPPPI and CBCL1.5-5, and written information about the study. Parents decided for themselves whether the father or mother would complete the questionnaire. Parents handed in the completed child health monitor questionnaire at the well child visit. The child health professional used the parent-completed KIPPPI as both a guide for the conversation with parents and as a tool for early detection of psychological and pedagogical problems. Based on the conversation with the parent and the completed child

health monitor questionnaire, the child health professional made a clinical decision whether a child is to be referred to a mental health care professional (e.g. psychologist) or whether a follow-up consultation is required. The child health professional registered the clinical decision on a separate registration form or in a digital medical record system. Although the KIPPPI can be scored, in this study the child health care professionals did not calculate scores since cutpoints were not empirically determined at the time.

Parents of 3655 (61.3%) children attended the well-child visit. The remaining parents (38.7%) neither attended the well-child visit nor completed the child health monitor questionnaire. Of those parents that did attend the well-child visit, 3164 (86.6%) had completed the child health monitor questionnaire. Children were excluded from the analyses if they were under treatment by a mental health professional at the time of inclusion (n=1; 0.03%), or if the KIPPPI contained more than 25% missing items on the KIPPPI scales (n=431, 13.6%). After exclusion, a study population of 2732 (86.3%) children was eligible for this study. The CBCL1.5-5 [16] was also included in the child health monitor questionnaire but only for research purposes (i.e. evaluating the concurrent validity and screening accuracy of the KIPPPI). Parents of 2016 (55.2%) children, who attended the well-child visit, also completed the CBCL1.5-5 in addition to the KIPPPI.

After the well-child visit, 225 parents were sent another copy of the KIPPPI to assess the test-retest reliability of the KIPPPI. The parents of 90 (40.0%) children returned the KIPPPI. The range of the period between completion of questionnaires was 5-78 days (mean=38.6, SD=17.5).

Mean child age was 24.1 months (SD=1.2), 47.7% were girls, and 72.1% of the children had a native Dutch ethnic background [17]. Mean age of the mother was 33.5 years (SD=4.8) and mean age of the father was 36.2 years (SD=5.5). In 92.6% of the cases, the mother or both parents were the respondent(s). See Table 3.1 for information on demographic characteristics of the study population.

Measures

The KIPPPI has 67 items regarding psychosocial child problems, which might be possible pedagogical challenges for the parents. The child health professional discusses the items with high scores (indicating a problem) with parents and assesses whether the difficulties stem from a problem in the child (i.e. psychosocial), or the parent (pedagogical), or the parent-child

Table 3.1 Characteristics of the study population, N=2732

	% of participants	Mean (SD)
Characteristics mother		
Age (years)		33.5 (4.8)
Born in the Netherlands	77.9	
Characteristics father		
Age (years)		36.2 (5.5)
Born in the Netherlands	76.6	
Characteristics children		
Age (months)		24.1 (1.2)
Gender (girls)	47.7	
Ethnic background (Dutch)[1]	72.1	
Family characteristics		
Two-parent household	88.6	
One-child family	40.8	
Respondent (mother or both parents)	92.6	

[1] A child is considered Dutch when both parents were born in the Netherlands.

interaction. The KIPPPI consist of a Wellbeing scale (31 items), Competence scale (25 items) and an Autonomy scale (11 items). The response options range from 0 ('(almost) never') to 3 ('(almost) always'), or reversed if the item is positively formulated. The KIPPPI Total score is the sum of the scale scores. Responses were summed for each scale and missing values on the KIPPPI items were imputed with within scale person-means [18]. High scores on the KIPPPI are less favourable. The possible score range of the KIPPPI Total score is 0-201, of the Wellbeing scale is 0-93, of the Competence scale 0-75, of the Autonomy scale 0-33. Wellbeing consists of five subscales that measure difficulties of the child with eating/drinking (4 items) and sleeping (3 items) and whether the child shows problems with activity (5 items), mood (5 items) and behaviour (14 items). The Competence scale consists of four subscales that measure unfavourable child cognitive development (4 items) and whether the child shows problems with language (4 items), play (3 items) and contact (14 items). The Autonomy scale consists of three subscales and measure whether the child has problems with toilet training (4 items), motor skills (3 items) and independence (4 items). See Table 3.2 for an overview of the (sub)scales and item examples. Additionally, the KIPPPI contains six additional items regarding the child's physical health, each with three response options ('good', 'average, 'bad'/ 'never', 'sometimes', 'often'), with a possible score range of 0-12. The physical health scale does not add to the KIPPPI Total score.

Table 3.2 Overview of the (sub)scales of the KIPPPI and item examples[1] with response options

Scale	Subscale	Item example	Response options
Wellbeing			
	Eating/drinking	My toddler does not like certain food or drinks	(almost) never, sometimes, often, (almost) always
	Sleeping	My toddler has nightmares	idem
	Activity	My toddler is overactive	idem
	Mood	My toddler is nervous, tense	idem
	Behaviour	My toddler is bad tempered	idem
Competence			
	Cognitive development	My toddler is easily persuaded to start a new activity	(almost) always, often, sometimes, (almost) never
	Language	My toddler speaks in sentences of 2 words or more	idem
	Play	My toddler likes playing games (e.g. peekaboo)	idem
	Contact	My toddler has difficulty adjusting	(almost) never, sometimes, often, (almost) always
Autonomy			
	Toilet training	My toddler wets his/her pants or diaper	(almost) never, sometimes, often, (almost) always
	Motor skills	My toddler bumps into things or falls	idem
	Indepence	My toddler tries to repair something that is broken	(almost) always, often, sometimes, (almost) never

[1] The KIPPPI is a Dutch questionnaire and for the purpose of this article some items are translate.

In addition to the KIPPPI, parents completed the CBCL1.5-5 in order to evaluate the concurrent validity and the screening accuracy of the KIPPPI. The well-validated [16] 100-item CBCL1.5-5 is designed for children aged 18-months to 5-years and has two domains (Internalising and Externalising) and provides a Total Problem score. Answers are given on a 3-point scale ('not true', 'somewhat or sometimes true' and 'very true or often true').

Analyses

Score distribution

Score distribution was evaluated by assessing the mean scale scores and standard deviations, and the presence of floor and ceiling effects (i.e. >15% of the respondents have the minimal and/or maximal score) [19]. Independent t-tests were performed to test the differences in mean KIPPPI scores between subgroups for child gender and ethnicity.

Reliability

Cronbach's alpha was used to evaluate the internal consistency of the KIPPPI-Total score, Wellbeing, Competence and Autonomy scales and their subscales. An alpha of 0.70 or higher was considered acceptable [20]. Test-retest reliability of the KIPPPI-scales was assessed with the Intraclass Correlation Coefficients (ICC), using a two-way random effect model with absolute agreement. An ICC of 0.70 or higher is considered to indicate acceptable test-retest reliability [19].

Validity

Concurrent validity was evaluated by assessing the Pearson correlation coefficients between KIPPPI and CBCL1.5-5 scale scores. Concurrent validity is hypothesized to be expressed in: large positive correlations between (a) KIPPPI-Total score, (b) KIPPPI Wellbeing and CBCL1.5-5 Internalising, Externalising and Total Problem scores, since the content of the items of the KIPPPI Total score and KIPPPI Wellbeing scale most resemble the items of the CBCL1.5-5. Furthermore we hypothesized there would be small to medium positive correlations between (c) KIPPPI Competence scale, (d) KIPPPI Autonomy scale and CBCL1.5-5 Internalising, Externalising and Total Problem scores, because the content of the items of the KIPPPI Competence scale and KIPPPI Autonomy scale have less overlap with items on the CBCL1.5-5. A correlation of 0.10 is considered small, 0.30 is considered medium and >0.50 is considered large [21].

Discriminative validity of the KIPPPI was evaluated by the ability of the KIPPPI to discriminate between a subgroup of parents who did and did not report being worried about their child's upbringing. We hypothesized that discriminative validity will be reflected in less favourable KIPPPI scores for children of parents who are worried about their child [22]. Regression analyses were used to evaluate the relationship between parental worry as independent variable and KIPPPI (scale) scores as dependent variable, corrected for confounding effects of child's gender and ethnicity. We hypothesized that parental worry is a significant predictor of KIPPPI scores. Effect sizes were defined as $Cohen's\ d = |\,[mean_{not\ worried} - mean_{worried}]/SD_{worried}\,|$ [21]; $0.20 \leq d < 0.50$ indicates a small effect, $0.50 \leq d < 0.80$ indicates a medium effect and $d \geq 0.80$ indicates a large effect.

Screening accuracy

Screening accuracy for the KIPPPI Total scores and scores on Wellbeing, Competence and Autonomy scales was evaluated against the CBCL1.5-5 as a golden standard (i.e. Total Problem

score in the clinical range), by calculating the Receiver Operating Characteristic (ROC) curves. The ROC curve is a plot of sensitivity as a function of 1-specificity for all possible cutpoints. The greater the area under the curve (AUC), the more discriminative the KIPPPI scores are. An AUC greater than 0.90 indicates high accuracy, AUC of 0.70-0.90 indicates moderate accuracy, 0.50-0.70 low accuracy, and 0.50 chance level accuracy [23].

A method to determine the optimal cutpoint for a test is calculating the Youden's index, which is defined as the maximum vertical distance between the ROC curve and the diagonal or chance line and is calculated as *Youden's index=sensitivity+specificity-1*. Screening accuracy for various cutpoints was evaluated by sensitivity, specificity, positive and negative likelihood ratio's (LHR^+ and LHR^-) and diagnostic odds ratio (OR).

Sensitivity is the proportion of true positives that are correctly identified by the test; specificity is the proportion of true negatives that are correctly identified by the test. In clinical practice, however, the test result is all that is known, knowledge whether or not someone is correctly classified is not available.

To overcome this problem, likelihood ratios can be calculated. LHR^+ is the ratio of the probability of a positive test result if the outcome is positive (true positive) to the probability of a positive test result if the outcome is negative (false positive); $LHR^+=(sensitvitiy/1\text{-}specificity)$. LHR^- is the ratio of the probability of a negative test result if the outcome is positive (false negative) to the probability of a negative test result if the outcome is negative (true negative); $LHR^-=(1\text{-}sensitivity/specificity)$. Tests with high screening accuracy have LHR^+ greater than 7 and LHR^- smaller than 0.30 [24].

The diagnostic odds ratio of a test is the ratio of the odds of a positive test result when having the 'disorder' relative to the odds of a positive test result when not having the 'disorder' and can be calculated as $(sensitivity*specificity)/((1\text{-}sensitivity)*(1\text{-}specificity))=LHR^+/LHR^-$. The values of OR ranges from zero to infinity, with higher values indicating better discriminatory test performance. Potentially useful tests tend to have diagnostic odds ratios well above 20 [24]. A value of 1 means that a test does not discriminate between people with and people without the 'disorder'. Values lower than 1 indicate improper test interpretation, meaning more negative tests among the people with the 'disorder' [25]. AUC, Youden's index, sensitivity, specificity, LHR^+, LHR^- and OR are independent of the prevalence of the 'disorder'.

We expected large AUCs for the KIPPPI total score and for the Wellbeing scale since the content of the items of these scales most resemble the items of the CBCL1.5-5 Total Problem

score. Whereas, we expected the AUCs of the Competence scale and Autonomy scale to be small (i.e. closer to 0.50) since the content of items of the Competence scale and Autonomy scale are less reflected in the CBCL1.5-5 items.

Clinical application

To evaluate the clinical application of the KIPPPI, registration data from the child health professionals (i.e. the clinical decision) was combined with the KIPPPI data from the parents.

The clinical application of the KIPPPI was explored by evaluating the relation between KIPPPI scores and the clinical decision of the child health professional. We hypothesized that the clinical decision of the child health professional to refer to another mental health professional or request a follow-up consultation, predicts higher KIPPPI scores, as high KIPPPI scores are expected to be indicative of problems. The data is hierarchical in nature since the child health professionals assessed more than one child, which makes (part of) the observations dependent on each other. Because the observations are not independent on each other, a multilevel regression analyses was used to evaluate the relation between the clinical decision as independent variable and the KIPPPI (scale-)scores as dependent variable, corrected for confounding effects of child's gender and ethnicity.

In this study we were able to combine 1448 (53.0%) of the parent-completed KIPPPI questionnaires with the clinical decision data registered by child health professionals. Combined data of 1284 (47.0%) children were lacking due to missing patient-codes. Significant differences ($p<0.05$) between the group with complete data and the group with incomplete data were found for the age of the child, ethnicity of the child and country of birth of the father and not for child gender, country of birth of the mother, age of the parents, family composition, person who completed the questionnaire and mean KIPPPI scores. Effect sizes of the significant differences between the group with complete data en the group with incomplete data, however, were very small (child age $d=0.16$, ethnicity $d=0.05$ and father country of birth $d=0.04$) and indicate that the data may be interpreted as 'missing at random'.

Multilevel regression analyses were performed in SAS software version 9.13 (SAS Institute Inc., 2009). All other analyses were performed in SPSS 19.0 (SPSS Inc. 2010).

RESULTS

Score distribution

Mean scale scores for the total sample and in subgroups by child's gender and ethnic background are presented in Table 3.3. Compared to girls, boys had significantly ($p<0.05$) higher mean KIPPPI Total scores and higher mean scores on the scales Wellbeing, Competence and Autonomy, and on the subscales Activity, Behaviour, Language, Play, Toilet training, Motor skills and Independence. Compared to Native children, non-native children had significantly ($p<0.05$) higher mean KIPPPI Total scores and higher mean scores on the scales Wellbeing and Competence, and on the subscales Sleeping, Activity, Mood, Behaviour, Cognitive development, Language, Play, Contact and Independence. Non-native children had significantly ($p<0.05$) lower scores on the KIPPPI subscale Toilet training. See Table 3.3.

Floor effects were present for seven subscales: Physical health, Eating/Drinking, Sleeping, Mood, Language, Play and Motor Skills. Ceiling effects were absent (Table 3.3).

Reliability

Internal consistency was 0.88 for the KIPPPI Total score; 0.86 for the Wellbeing scale; 0.83 for the Competence scale; and 0.58 for the Autonomy scale (Table 3.3). The internal consistency of the subscales is presented in Table 3.3. Only the subscales Eating/Drinking, Behaviour, Language and Contact had Cronbach alpha's greater than 0.70.

Test-retest reliability was 0.80 for the KIPPPI Total score; 0.76 for the Wellbeing scale; 0.73 for the Competence scale; and 0.60 for the Autonomy scale (Table 3.3). The test-retest reliability of the subscales is presented in Table 3.3. Only the subscales Physical health, Sleeping, Activity, Behaviour, Language, Contact, Motor skills and Independence had ICCs greater than 0.70.

Validity

Concurrent validity

As hypothesized, positive correlations were found between the KIPPPI Total score and the CBCL1.5-5 scores for Internalising ($r=0.60$), Externalising ($r=0.63$) and Total Problem score ($r=0.68$). The KIPPPI Wellbeing scale was positively correlated with the CBCL1.5-5 scores for Internalising ($r=0.55$), Externalising ($r=0.74$) and Total Problem score ($r=0.72$).

Table 3.3 Score distributions, internal consistency and test-retest reliability of the KIPPPI-scales, N_{total}=2732

KIPPPI scale (# items)	Mean (SD) Total	Mean (SD) Boys (N=1409)	Mean (SD) Girls (N=1304)	Mean (SD) Native (N=1969)	Mean (SD) Non-native (N=763)	%min[1] Total	%max[1] Total	Cronbach's α Total	Test-retest ICC[2] N=90
KIPPPI TOTAL (67)	41.7 (14.5)	43.6[a] (14.7)	39.6[a] (14.1)	40.5[b] (14.0)	44.8[b] (15.5)	0.0	0.0	0.88	0.80
Physical health (6)	1.0 (1.2)	1.1 (1.3)	1.0 (1.2)	1.0 (1.2)	1.1 (1.3)	43.7	0.0	0.38	0.87
Wellbeing (31)	17.0 (8.4)	17.6[a] (8.6)	16.3[a] (8.2)	16.4[b] (8.2)	18.6[b] (8.7)	0.0	0.0	0.86	0.76
Eating/Drinking (4)	2.4 (2.0)	2.4 (2.0)	2.3 (2.0)	2.3 (1.9)	2.4 (2.1)	21.3	0.0	0.70	0.63
Sleeping (3)	1.3 (1.4)	1.3 (1.4)	1.3 (1.4)	1.3[b] (1.4)	1.5[b] (1.4)	35.7	0.0	0.59	0.82
Activity (5)	3.5 (2.3)	3.7[a] (2.4)	3.2[a] (2.2)	3.4[b] (2.4)	3.6[b] (2.3)	9.3	0.0	0.67	0.78
Mood (5)	1.1 (1.6)	1.1 (1.5)	1.2 (1.7)	0.9[b] (1.3)	1.7[b] (1.9)	46.0	0.0	0.60	0.45
Behaviour (14)	8.7 (4.7)	9.0[a] (4.8)	8.3[a] (4.5)	8.4[b] (4.6)	9.3[b] (4.7)	2.0	0.0	0.82	0.80
Competence (25)	11.6 (7.1)	12.2[a] (7.3)	11.0[a] (6.8)	11.0[b] (6.7)	13.1[b] (7.6)	1.1	0.0	0.83	0.73
Cognitive development (4)	2.0 (1.5)	2.0 (1.5)	1.9 (1.5)	1.9[b] (1.4)	2.1[b] (1.6)	13.5	0.0	0.61	0.46
Language (4)	2.6 (2.8)	3.1[a] (3.0)	2.1[a] (2.4)	2.3[b] (2.6)	3.4[b] (3.0)	30.4	0.8	0.79	0.79
Play (3)	0.8 (1.1)	0.8[a] (1.2)	0.7[a] (1.1)	0.7[b] (1.0)	0.9[b] (1.2)	56.2	0.0	0.60	0.62
Contact (14)	5.8 (4.5)	5.8 (4.5)	5.8 (4.5)	5.6[b] (4.3)	6.3[b] (4.8)	8.8	0.0	0.82	0.71
Autonomy (11)	13.1 (3.3)	13.8[a] (3.1)	12.4[a] (3.4)	13.2 (3.1)	13.0 (3.6)	0.0	0.0	0.58	0.60
Toilet training (4)	7.9 (2.0)	8.3[a] (1.8)	7.5[a] (2.2)	8.0[b] (1.9)	7.7[b] (2.4)	1.1	1.8	0.59	0.60
Motor skills (3)	1.1 (0.8)	1.2[a] (0.8)	1.1[a] (0.8)	1.1 (0.8)	1.1 (0.9)	17.7	0.0	0.18	0.70
Independence (4)	4.1 (1.9)	4.3[a] (1.9)	3.8[a] (1.9)	4.0[b] (1.8)	4.2[b] (2.0)	0.0	0.0	0.49	0.74

[1] % of respondents with the lowest (min) and highest (max) BITSEA scale score (ceiling/floor).

[2] Test-retest Intraclass Correlation Coefficients are all significant, p<0.01.

a = significant difference in mean BITSEA scores between boys and girls, p<0.05.

b = significant difference in mean BITSEA scores between native and non-native children, p<0.05.

The KIPPPI Competence scale was positively correlated with the CBCL1.5-5 scores for Internalising (r=0.48), Externalising (r=0.32) and Total Problem score (r=0.43). The KIPPPI Autonomy scale was positively correlated with the CBCL1.5-5 scores for Internlising (r=0.18), Externalising (r=0.23) and Total Problem score (r=0.23). All these correlations were significant, p<0.01. See Table 3.4 for the concurrent validity of the subscales.

Discriminative validity

KIPPPI scores of 2109 (77.2%) children of parents who did not report to be worried about their child's upbringing were compared to KIPPPI scores of 604 (22.1%) children of parents who did report to be worried (percentages do not sum to 100% because of missing values). All regression coefficients were significant (p<0.01) and positive: KIPPPI Total, B=11.87; Wellbeing scale B=7.38; Competence scale, B=3.40; and Autonomy scale, B=1.10. See Table 3.5. The effect sizes of the differences in mean KIPPPI scores between parents that did and did not report to be worried about their child's upbringing ranged from large to small: KIPPPI Total score, d=0.81; for the Wellbeing scale, d=0.85; for the Competence scale, d=0.46; and

Table 3.4 Concurrent validity (Pearson correlation coefficients) between KIPPPI scales and CBCL1.5-5 Internalising, Externalising and Total Problem score, N=2016

	CBCL scales		
KIPPPI scales	Internalising	Externalising	Total Probem
KIPPPI TOTAL	0.60	0.63	0.68
Physical health	0.28	0.20	0.27
Wellbeing	0.55	0.74	0.72
Eating/Drinking	0.26	0.21	0.26
Sleeping	0.27	0.26	0.37
Activity	0.31	0.61	0.51
Mood	0.49	0.36	0.46
Behaviour	0.49	0.73	0.67
Competence	0.48	0.32	0.43
Cognitive development	0.23	0.17	0.22
Language	0.18	0.17	0.19
Play	0.28	0.33	0.33
Contact	0.50	0.26	0.40
Autonomy	0.18	0.23	0.23
Toilet training	0.04	0.10	0.07
Motor skills	0.23	0.32	0.33
Independence	0.16	0.15	0.17

Note: Underlined correlation is non-significant (p>0.05), all other correlations are significant, p<0.01.

Table 3.5 Discriminative ability of the KIPPPI between subgroups differing in parental worries about the child's upbringing

	Parental worries		Beta[1]	Effect size[2]
	Mean (SD)			
	Not worried N=2109	Worried N=604		
KIPPPI TOTAL	39.0 (13.1)	51.2 (15.0)	11.87	0.81[a]
Physical health	0.9 (1.1)	1.4 (1.4)	0.49	0.36[c]
Wellbeing	15.3 (7.4)	22.8 (8.8)	7.38	0.85[a]
Eating/drinking	2.1 (1.9)	3.1 (2.1)	0.94	0.48[c]
Sleeping	1.2 (1.3)	1.8 (1.6)	0.62	0.38[c]
Activity	3.2 (2.2)	4.5 (2.6)	1.28	0.50[b]
Mood	1.0 (1.4)	1.7 (1.9)	0.69	0.37[c]
Behaviour	7.8 (4.1)	11.7 (5.1)	3.84	0.76[b]
Competence	10.8 (6.6)	14.4 (7.8)	3.40	0.46[c]
Cognitive development	1.9 (1.4)	2.3 (1.6)	0.39	0.25[c]
Language	2.4 (2.6)	3.3 (3.1)	0.78	0.29[c]
Play	0.7 (1.0)	1.1 (1.3)	0.46	0.31[c]
Contact	5.4 (4.3)	7.2 (4.8)	1.78	0.38[c]
Autonomy	12.9 (3.2)	14.0 (3.3)	1.10	0.33[c]
Toilet training	7.8 (2.0)	8.2 (2.1)	0.33	0.19[c]
Motor skills	1.1 (0.8)	1.4 (0.9)	0.34	0.33[c]
Independence	4.0 (1.9)	4.4 (1.9)	0.42	0.21[c]

[1] Unstandardized Beta's are corrected for confounding effects of child's gender and ethnicity and significant, $p<0.01$.
[2] Difference of the means divided by SD in the subgroup 'worried'.
[a] indicates a large effect ($d\geq0.80$).
[b] indicates a medium effect ($0.50\leq d<0.80$).
[c] indicates a small effect ($0.20\leq d<0.50$).

for the Autonomy scale, d=0.33. See Table 3.5 also for the effect sizes of the subscales. The subscales Activity and Behaviour had medium effect sizes whereas the effect size for all other subscales was small.

Screening accuracy

ROC curves of the KIPPPI Total score, Wellbeing scale, Competence scale and Autonomy scale are presented in Figure 3.1. In Table 3.6 AUC and sensitivity, specificity, LHR^+, LHR^-, OR and Youden's index are presented for a range of KIPPPI cutpoints. The AUC for the KIPPPI Total score was 0.92 and for the Wellbeing scale 0.93. The AUC for the Competence and Autonomy scale were lower; respectively 0.81 and 0.60.

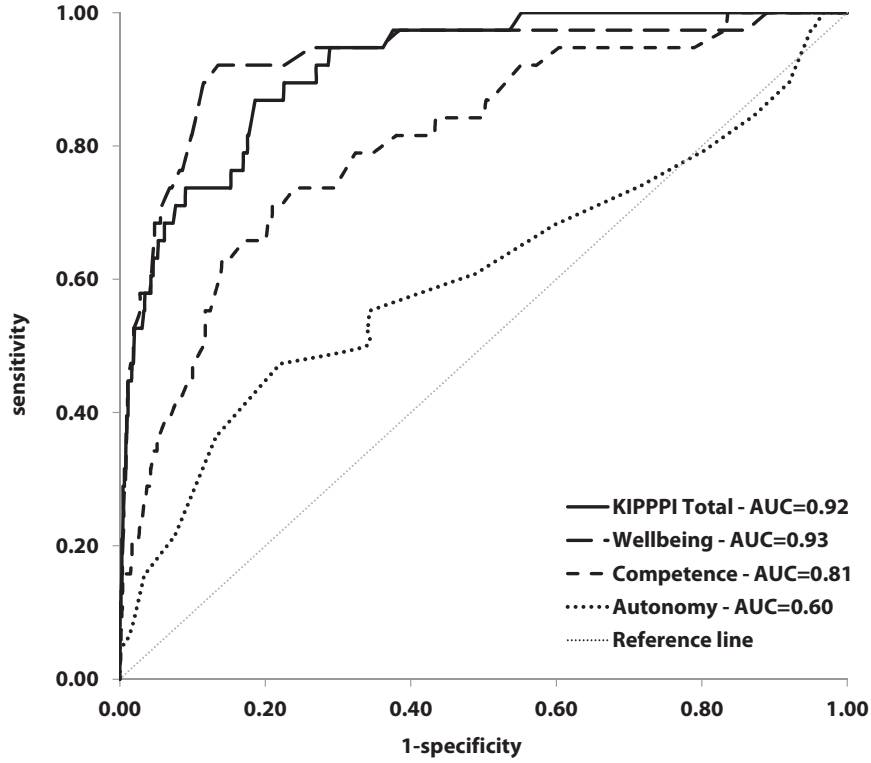


Figure 3.1 Receiver Operating Characteristic curve for KIPPPI scales Wellbeing, Competence, Autonomy and KIPPPI Total score, relative to CBCL1.5-5 Total Problem score in the clinical range. AUC=area under the curve.

Clinical application

Child health professionals referred 149 (10.0%) children for further evaluation or requested a follow-up consultation. All regression coefficients were significant ($p<0.05$) and positive: KIPPPI Total score, B=11.00; Wellbeing scale, B=4.95; Competence scale, B=4.73; and Autonomy scale, B=1.31. The effect sizes of the differences in mean KIPPPI scores between children that did and did not need referral or a follow-up consultation were for the KIPPPI Total score d=0.65; for the Wellbeing scale d=0.50; for the Competence scale d=0.61; and for the Autonomy scale d=0.32. See Table 3.7 also for the effect sizes of the subscales. The subscale Language had a large effect size, the subscale Behaviour had a medium effect sizes and all other subscales had small effect sizes.

Table 3.6 Area Under the Curve and sensitivity, specificity, liklihood ratio's and Youden's index for a range of KIPPPI scores relative to a clinical CBCL Total Problem score

KIPPPI Total							KIPPPI Wellbeing							KIPPPI Competence							KIPPPI Autonomy						
AUC=0.92 (95% CI=0.88-0.96)							AUC=0.93 (95% CI=0.88-0.98)							AUC=0.81 (95% CI=0.74-0.88)							AUC=0.60 (95% CI=0.50-0.71)						
score	sens	spec	LHR^+	LHR^-	OR	J	score	sens	spec	LHR^+	LHR^-	OR	J	score	sens	spec	LHR^+	LHR^-	OR	J	score	sens	spec	LHR^+	LHR^-	OR	J
50	0.92	0.73	3.41	0.11	31.09	0.65	22	0.95	0.74	3.65	0.07	54.08	0.68	10	0.92	0.45	1.67	0.18	9.41	0.37	9	0.89	0.08	0.97	1.38	0.70	-0.02
51	0.89	0.76	3.71	0.14	25.62	0.65	23	0.92	0.78	4.18	0.10	40.77	0.70	11	0.87	0.50	1.74	0.26	6.69	0.37	10	0.84	0.13	0.97	1.23	0.78	-0.03
52	0.89	0.77	3.87	0.14	27.09	0.67	24	0.92	0.81	4.84	0.10	49.03	0.73	12	0.84	0.57	1.95	0.28	6.96	0.41	11	0.79	0.20	0.99	1.05	0.94	-0.01
53	0.87	0.80	4.35	0.16	26.77	0.67	25	0.92	0.84	5.75	0.10	60.38	0.76	13	0.82	0.62	2.16	0.29	7.43	0.44	12	0.74	0.29	1.04	0.90	1.16	0.03
54	**0.87**	**0.81**	**4.58**	**0.16**	**28.53**	**0.68**	**26**	**0.92**	**0.87**	**7.08**	**0.09**	**76.96**	**0.79**	14	0.79	0.68	2.47	0.31	7.99	0.47	13	0.68	0.40	1.13	0.80	1.42	0.08
55	0.82	0.82	4.56	0.22	20.75	0.64	27	0.89	0.88	7.42	0.13	59.33	0.78	15	0.74	0.72	2.64	0.36	7.32	0.46	14	0.61	0.52	1.27	0.75	1.69	0.12
56	0.76	0.85	5.07	0.28	17.94	0.61	28	0.82	0.90	8.20	0.20	41.00	0.72	16	0.74	0.76	3.08	0.34	9.01	0.50	15	0.55	0.66	1.62	0.68	2.37	0.21
57	0.74	0.87	5.69	0.30	19.05	0.61	29	0.76	0.92	9.50	0.26	36.42	0.68	**17**	**0.71**	**0.79**	**3.38**	**0.37**	**9.21**	**0.50**	**16**	**0.47**	**0.78**	**2.14**	**0.68**	**3.14**	**0.25**
58	0.74	0.89	6.73	0.29	23.03	0.62	30	0.74	0.93	10.57	0.28	37.81	0.67	18	0.66	0.83	3.88	0.41	9.48	0.49	17	0.37	0.87	2.85	0.72	3.93	0.23
59	0.74	0.89	6.73	0.29	23.03	0.63	31	0.71	0.94	11.83	0.31	38.36	0.65	19	0.63	0.86	4.50	0.43	10.46	0.49	18	0.21	0.93	3.00	0.85	3.53	0.14
60	0.74	0.90	7.40	0.29	25.62	0.64	32	0.68	0.95	13.60	0.34	40.38	0.64	20	0.55	0.88	4.58	0.51	8.96	0.44	19	0.16	0.97	5.33	0.87	6.16	0.12
61	0.74	0.91	8.22	0.29	28.78	0.65	33	0.58	0.96	14.50	0.44	33.14	0.54	21	0.47	0.90	4.70	0.59	7.98	0.37	20	0.08	0.98	4.00	0.94	4.26	0.06
62	0.71	0.92	8.88	0.32	28.16	0.63	34	0.58	0.97	19.33	0.43	44.65	0.55	22	0.42	0.92	5.25	0.63	8.33	0.34	21	0.05	0.99	5.00	0.96	5.21	0.05
63	0.68	0.93	9.71	0.34	28.23	0.61	35	0.58	0.97	19.33	0.43	44.65	0.55	23	0.39	0.94	6.50	0.65	10.02	0.33	22	0.05	1.00	.	0.95	.	0.05

Note: AUC=area under the curve; 95%CI=95% confidence interval; sens=sensitivity; spec=specificity; LHR^+=liklihood ratio positive test; LHR^-=liklihood ratio negative test; OR=diagnostic odds ratio; J=Youden's index. All AUCs were significant ($p<0.01$). Scores with the highest unroundend Youden's index are indicated in bold.

Table 3.7 Clinical application of the KIPPPI; relation between KIPPPI scores and the decision by the child health professional to refer to a specialist and/or request a follow-up consultation

KIPPPI scales	Referral or follow-up decision		Beta[1]	Effect size[2]
	Mean (SD)			
	Not referred N=1335	Referred N=149		
KIPPPI TOTAL	41.8 (14.0)	53.0 (17.2)	11.00	0.65[b]
Physical health	1.0 (1.2)	1.3 (1.6)	0.33	0.19[c]
Wellbeing	16.1 (7.9)	21.3 (10.5)	4.95	0.50[b]
Eating/drinking	2.2 (2.0)	2.7 (2.2)	0.57	0.23[c]
Sleeping	1.3 (1.4)	1.8 (1.7)	0.51	0.29[c]
Activity	3.4 (2.3)	4.2 (2.6)	0.77	0.31[c]
Mood	1.0 (1.3)	1.9 (2.2)	0.86	0.41[c]
Behaviour	8.4 (4.6)	11.4 (6.0)	3.21	0.50[b]
Competence	12.9 (7.3)	18.0 (8.4)	4.73	0.61[b]
Cognitive development	2.0 (1.5)	2.4 (1.7)	0.58	0.24[c]
Language	2.4 (2.6)	5.4 (3.6)	2.73	0.83[a]
Play	0.7 (1.0)	0.9 (1.3)	0.24	0.15[c]
Contact	7.5 (4.9)	9.2 (5.5)	1.69	0.31[c]
Autonomy	13.0 (3.4)	14.2 (3.7)	1.31	0.32[c]
Toilet training	8.0 (2.0)	8.7 (2.0)	0.68	0.35[c]
Motor skills	1.1 (0.8)	1.4 (1.0)	0.28	0.30[c]
Independence	4.0 (1.9)	4.5 (1.9)	0.43	0.26[c]

[1] Unstandardized Beta's are corrected for confounding effects of child's gender and ethnicity and significant, $p<0.05$.
[2] Difference of the means divided by SD in the subgroup 'intervention needed' and 'referred'.
[a] indicates a large effect ($d\geq0.80$).
[b] indicates a medium effect ($0.50\leq d<0.80$).
[c] indicates a small effect ($0.20\leq d<0.50$).

DISCUSSION

The present study evaluated the psychometric properties of the KIPPPI, a Dutch instrument that was developed to measure psychological and pedagogical problems in 2-year-olds, in a large community sample. The score distribution and the following psychometric properties of the KIPPPI were determined: internal consistency, test-retest reliability, concurrent validity, discriminative validity and screening accuracy. Additionally we also evaluated the clinical application of the KIPPPI.

Score distribution

The KIPPPI scales and KIPPPI Total score showed no floor or ceiling effects. Floor effects were present, however, for the following subscales: Physical health, Eating/Drinking, Sleeping, Mood, Language, Play and Motor Skills. This means that changes within toddlers with low scores for these subscales cannot be measured and that there is less differentiation possible between children with low KIPPPI scores (i.e. few psychosocial problems) [19]. The mean KIPPPI Total score and KIPPPI scale scores were less favourable for boys compared to girls and for non-native children compared to native children. There was, however, no difference in mean score on the Autonomy scale between native and non-native children. These findings are in line with previous studies that report boys experience psychosocial problems more often than girls [26] and that psychosocial problems are more frequently reported in immigrant children compared to native children [27,28].

Reliability

Internal consistency for the KIPPPI Total score, Wellbeing scale and Competence scale was adequate (>0.70) whereas the internal consistency for the Autonomy scale was marginal (i.e. 0.58). Lower internal consistency for the Autonomy scale might be explained by the inclusion of some items that assess behaviours that may not be expected to co-occur, for example: "Runs and climbs" and "Tries to repair something that is broken".

The 5-78 day (mean=38.6, SD=17.5) test-retest reliability was adequate (>0.70) for the KIPPPI Total score, Wellbeing scale and Competence scale and was marginal (i.e. 0.60) for the Autonomy scale. These results mean that, assuming that no real changes in psychosocial problems occur, the KIPPPI Total score, Wellbeing scale and Competence scale provide stable outcome measures over time.

Validity

As hypothesized, the KIPPPI showed good concurrent validity: the KIPPPI Total score and Wellbeing scale had large positive correlations with CBCL1.5-5 Internalising, Externalising and Total Problem scores. Also, as hypothesized the Competence scale and Autonomy scale had a small to medium positive correlation with CBCL1.5-5 Internalising, Externalising and Total Problem scores.

The KIPPPI Total score and scale scores were able to distinguish between parents who reported being worried about their child's upbringing and parents who did not report being worried. This indicates that scores were less favourable for children of parents who were worried, compared to parents that were not worried. The difference between these subgroups in mean KIPPPI Total score and mean scores on the Wellbeing scale was large ($d \geq 0.80$) However, the difference in mean scores on the Competence scale and Autonomy scale was small ($0.20 \leq d < 0.50$). These results indicate good discriminative validity for the KIPPPI Total score and Wellbeing scale.

Screening accuracy

The KIPPPI showed large Areas Under the Curve (>0.90) for the KIPPPI Total score and Wellbeing scale and indicates that these scores have high accuracy in discriminating between children with psychosocial problems and children without psychosocial problems. The Competence scale AUC showed moderate accuracy (AUC=0.81) and Autonomy scale AUC showed low accuracy (AUC=0.60). The KIPPPI Total score and Wellbeing scale are better able to discriminate between children with psychosocial problems and children without psychosocial problems, compared to the Competence scale and Autonomy scale.

Clinical application

KIPPPI Total score and scale scores were positive and significantly associated with child health professional's decision whether or not a follow-up consultation or referral was required. The difference between children who were referred and children we were not referred in mean KIPPPI Total score and mean scores on the Wellbeing scale and Competence scale was medium ($0.50 \leq d < 0.80$). However the difference between these subgroups in mean scores on the Autonomy scale was small ($0.20 \leq d < 0.50$). These results indicate that scores were less favourable for children who were referred or asked back for a follow-up consultation, compared to children who were not referred or asked back for a follow-up consultation.

Limitations and strengths

Our study has two main limitations. First, in the current study we have no data on the non-response group, because no information is available on parents who did not attend the well-child visit. Therefore, some care should be taken with generalizing these results to the total

population. However, due to the diversity of our large study population, we do not expect that the characteristics of the non-response group are very different of that from the study population. In the Netherlands, participation of parents with their child in the preventive youth health care is free of charge, which makes the well-child visit easily accessible for all population groups: There is no dissimilarity in visiting frequency between native Dutch and non-native children and their parents [29].

Second, the report by parents introduces the proxy-problem: self-report by two-year-old children on their psychosocial problems is not possible, because children of this age lack the necessary language skills and the cognitive abilities to interpret the questions and they do not have a long-term view of events [30]. Although reports by parents do not provide first-hand information and answers might be clouded by how a parent interprets their child's behaviour, proxy by parents in this case might be a useful alternative [31].

A major strength of our study is the large and diverse sample. Additionally, the setting in which the respondents were invited to complete the KIPPPI, the daily practice of well-child-visit at the child health care centre, can be seen as either a strength or a limitation. We evaluated the psychometric properties in a setting in which the KIPPPI is used; however this specific setting might, on the other hand, hamper the generalizations of our results to other settings.

Conclusions

The psychometric properties of the KIPPPI are comparable to that of other early detection tools for preschool children [32]. Early detection instruments for psychosocial problems in infants and toddlers are scarce [33]. The Child Behavioral Checklist (CBCL1.5-5) [16] has good reliability and validity, but is too long to employ as an early detection tool in preventive child health care. The KIPPPI addresses both problem behaviour as well as competencies, but unlike the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) [34] and the Ages & Stages Questionnaire-Social-Emotional version (ASQ-SE) [35], the KIPPPI does not consist of items specifically for the early detection of autism spectrum disorders. The KIPPPI covers a wide range of psychological and pedagogical aspects of a child's development, which might make it appealing to use by a child health professional during the well-child visit.

We recommend future studies to evaluate the psychometric properties of the KIPPPI, also in a different sample and setting. The setting of this study was the daily practice of a well-child visit in an urban area, however, it would be good to replicate this study in a more rural area,

possibly outside the context of a well-child visit. Future studies may also wish to further investigate differences in KIPPPI psychometric properties for population subgroups (e.g. child gender and ethnic background). Furthermore, differences in screening accuracy and cutpoints for boys and girls might be explored, since these groups have different mean KIPPPI scores. Although the KIPPPI showed adequate screening accuracy relative to the CBCL1.5-5, we recommend further evaluation of the screening accuracy of the KIPPPI by including a clinical sample of children with a clinical diagnosis made by a (mental health) professional.

In conclusion, the results of our study support the reliability, validity and clinical application of the KIPPPI Total score, Wellbeing scale and Competence scale. Also, the screening accuracy of the KIPPPI Total score and Wellbeing scale were supported. The Autonomy scale needs further study.

ACKNOWLEDGMENTS

We would like to express our gratitude to Irene Hoekstra for her assistance in this study.

REFERENCES

1. American Academy of Pediatrics. Committee on children with disabilities: development surveillance and screening of infants and young children. (2001) Pediatrics 108: 192-196.
2. American Academy of Pediatrics. Identifying infants and young children with development disorders in the medical home: an algorithm for developmental surveillance and screening. (2006) Pediatrics 118: 405-420.
3. Hermans J, Ory F, Schrijvers G (Julius Centrum: Utrecht; 2005) Supporting development and parenting: sooner, faster and better: an advice for early detection and interventions in regard to developmental and parenting problems.
4. Klein Velderman M, Crone MR, Wiefferink CH, Reijneveld SA (2010) Identification and management of psychosocial problems among toddlers by preventive child health care professionals. European Journal of Public Health 20: 332-338.
5. Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP (2004) Identification and management of psychosocial problems among toddlers in Dutch preventive child health care. Archives of Pediatrics & Adolescent Medicine 158: 811-817.
6. Lavigne JV, Arend R, Rosenbaum D, Binns HJ, Christoffel KK, et al. (1998) Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry 37: 1246-1254.
7. Mesman J, Koot HM (2001) Early preschool predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. Journal of the American Academy of Child and Adolescent Psychiatry 40: 1029-1036.
8. Durlak JA, Wells AM (1998) Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. American Journal of Community Psychology 26: 775-802.
9. Elliot J, Prior M, Merrigan C, Ballinger K (2002) Evaluation of a community intervention programme for preschool behaviour problems. Journal of Paediatrics and Child Health 38: 41-50.
10. Verhulst FC, van der Ende J (1997) Factors associated with child mental health service use in the community. Journal of the American Academy of Child and Adolescent Psychiatry 36: 901-909.
11. Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP (2001) Identification and management of psychosocial problems by preventive child health care. Archives of Pediatrics & Adolescent Medicine 155: 462-469.
12. Glascoe FP (2005) Screening for developmental and behavioral problems. Mental Retardation and Developmental Disabilities Research Reviews 11: 173-179.
13. Sand N, Silverstein M, Glascoe FP, Gupta VB, Tonniges TP, et al. (2005) Pediatricians' reported practices regarding developmental screening: Do guidelines work? Do they help? Pediatrics 116: 174-179.

14. Sturner RA (1991) Parent questionnaires - Basic office equipment. Journal of Developmental and Behavioral Pediatrics 12: 51-54.

15. Kousemaker NPJ (1996) Zoeken, vinden, zorgen delen: de ontwikkeling van een praktijkparadigma voor onderkenning en pedagogische preventie van psychosociale problematiek in de Jeugdgezondheidszorg [Searching, finding, sharing care; the development of a practise paradigm for the descernment and pedagogical prevention of psychosocial problems in Youth Healthcare].

16. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington:VT: University of Vermont, Research Center for Children, Youth, and Families.

17. Definitions 'native' and 'immigrant' (2012) - http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?conceptid=37. [Statistics Netherlands].

18. Shrive FM, Stuart H, Quan H, Ghali WA (2006) Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. BMC Med Res Methodol 6: 57.

19. Terwee CB, Bot SDM, de Boer MR, van der Windt D, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 60: 34-42.

20. Nunnally JC, Bernstein IH (1994) Psychometric theory. New York: McGraw-Hill.

21. Cohen J (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erblaum Associates.

22. Glascoe FP (1997) Parents' concerns about children's development: Prescreening technique or screening test? Pediatrics 99: 522-528.

23. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

24. Fischer JE, Bachmann LM, Jaeschke R (2003) A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Medicine 29: 1043-1051.

25. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM (2003) The diagnostic odds ratio: a single indicator of test performance. Journal of Clinical Epidemiology 56: 1129-1135.

26. Kraemer S (2000) The fragile male. British Medical Journal 321:1609.

27. Reijneveld S, Harland P, Brugman E, Verhulst F, Verloove-Vanhorick S (2005) Psychosocial problems among immigrant and non-immigrant children - Ethnicity plays a role in their occurrence and identification. European Child and Adolescent Psychiatry 14: 145-152.

28. Vollebergh WAM, ten Have M, Dekovic M, Oosterwegel A, Pels T, et al. (2005) Mental health in immigrant children in the Netherlands. Social Psychiatry and Psychiatric Epidemiology 40: 489-496.

29. Parents positive about child health clinics (2005) - http://www.cbs.nl/nl-NL/menu/themas/vrije-tijd-cultuur/publicaties/artikelen/archief/2005/2005-1734-wm.htm. [Statistics Netherlands].

30. Vogels T, Verrips GHW, Verloove-Vanhorick SP, Fekkes M, Kamphuis RP, et al. (1998) Measuring health-related quality of life in children: the development of the TACQOL parent form. Quality of Life Research 7: 457-465.

31. Theunissen NCM, Vogels TGC, Koopman HM, Verrips GHW, Zwinderman KAH, et al. (1998) The proxy problem: child report versus parent report in health-related quality of life research. Quality of Life Research 7: 387-397.

32. Sosna T, Mastergeorge A (2005) Compedium of screening tools for early childhood social-emotional development.

33. Glascoe FP (2000) Early detection of developmental and behavioral problems. Pediatrics in Review 21: 272-280.

34. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.

35. Squires J (2002) Ages & Stages Questionnaire Social-Emotional: A parent-completed, child monitoring system for social-emotional behaviors. Baltimore: Paul H. Brookes Publishing Co., Inc.

Chapter 4

Screening accuracy and clinical application of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA)

Ingrid Kruizinga
Wilma Jansen
Cathelijne L Mieloo
Alice S. Carter
Hein Raat

Published as:
Kruizinga I, Jansen W, Mieloo CL, Carter AS, Raat H (2013) Screening Accuracy and Clinical Application of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). PLoS ONE 8(8): e72602. doi:10.1371/journal.pone.0072602

ABSTRACT

Background: The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) is a promising questionnaire for the early detection of psychosocial problems in toddlers. The screening accuracy and clinical application were evaluated.

Methods: In a community sample of 2-year-olds (N=2060), screening accuracy of the BITSEA Problem scale was examined regarding a clinical CBCL1.5-5 Total Problem score. For the total population and subgroups by child's gender and ethnicity Receiver Operating Characteristic (ROC) curves were calculated, and across a range of BITSEA Problem scores, sensitivity, specificity, likelihood ratios, diagnostic odds ratio and Youden's index. Clinical application of the BITSEA was examined by evaluating the relation between the scale scores and the clinical decision of the child health professional.

Results: The area under the ROC curve (95% confidence interval) of the Problem scale was 0.97(0.95-0.98), there were no significant differences between subgroups. The association between clinical decision and BITSEA Problem score (B=2.5) and Competence score (B=-0.7) was significant ($p<0.05$).

Conclusion: The results indicate that the BITSEA Problem scale has good discriminative power to differentiate children with and without psychosocial problems. Referred children had less favourable scores compared to children that were not referred. The BITSEA may be helpful in the early detection of psychosocial problems.

INTRODUCTION

Preventive child health care has always focused on the detection of physical conditions. Recently, its focus has been expanded to mental health issues [1], offering an opportunity for the early detection of psychosocial problems among preschool children. In the Netherlands, preventive child health care for young children is delivered through community well-child clinics that are free of charge and provide routine developmental assessment and vaccinations [2]. One approach for facilitating early recognition and identification of psychosocial problems is to use parent-completed questionnaires as part of routine primary care visits (i.e. well-child visits) [3]. Early detection instruments of psychosocial problems, intended for use in preventive child health care, should have adequate psychometric properties (i.e. reliability and validity), and should also be short, easy to administer, score, and interpret [4,5]. Furthermore, early detection instruments should be able to correctly discriminate children with and without psychosocial problems. Of course early detection will not be without errors, but should be as accurate as possible as to minimize the expenses associated with over-referrals and under-detection [4]. The identification of cutpoints and their concomitant accuracy is therefore important, since this enables child health professionals to determine how many responses that indicate psychosocial problems, reliably indicating the actual presence of psychosocial problems. When cutpoints are identified, indices for screening accuracy (e.g. sensitivity and specificity) can be established for the questionnaire.

In the setting of preventive child health care general early detection instruments are warranted, since the aim is to early detect a broad range of possible psychosocial problems. The Child Behavior Checklist 1.5-5 (CBCL1.5-5) [6] and Infant-Toddler Social and Emotional Assessment (ITSEA) [7,8] are early detection instruments that are well-validated and measure a broad range of psychosocial problems, and in the case of the ITSEA also delays in competencies. However both instruments are too extensive to apply in the context of well-child visits. The ITSEA has been reported to have adequate psychometric properties [7,8], and exists in a shorter version; the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). The BITSEA comprises 42 items that measure psychosocial problems as well as delays in the acquisition of competencies in toddlers (1-3 year olds). Previous studies have shown that the BITSEA has adequate reliability and validity [9-11].

Previous studies have evaluated the sensitivity and specificity of the BITSEA [9,12]. One study, performed in Turkey among a community sample of 462 children, examined the sensitivity and specificity of only the BITSEA Competence scale relative to children treated in a child

psychiatry outpatient clinic with an autism diagnosis. In this study, the sensitivity was 72%-93% and specificity was 76-85%, depending on the cutpoint chosen [12]. In another study, performed in the United States of America, the sensitivity and specificity was examined in a community sample of 1280 children. In this study, the BITSEA Problem scale had, relative to the CBCL1.5-5 a sensitivity of 93.2% and a specificity of 78.0%. The BITSEA Competence scale was examined relative to low ITSEA Competence, and had a sensitivity of 68.9% and a specificity of 95.1% [9]. BITSEA Problem scale cutpoints were chosen at scores of ≥75th percentile and for the BITSEA Competence scale cutpoints were chosen at scores of <15th percentile [13].

In the present study we aim to evaluate the screening accuracy more extensively than prior studies. The screening accuracy will be evaluated with multiple indices (i.e. area under the curve, sensitivity, specificity, likelihood ratios, diagnostic odds ratio and Youden's index) by calculating receiver operating characteristic (ROC) curves of the BITSEA Problem scale relative to a CBCL Total Problem score in the clinical range. In our study we present indices of screening accuracy for a range of BITSEA Problem scores, because different cutpoints might be chosen in different settings (e.g. clinical application versus epidemiological research). The screening accuracy of the BITSEA Competence scale was not evaluated with a reference group of children with a CBCL Total Problem score in the clinical range, since the CBCL Total Problem score does not measure competencies. Previous studies showed differences in mean BITSEA scores between boys and girls (with boys scoring less favourably) [9-11] and between native and non-native children (with non-native children scoring less favourably) [11], therefore the screening accuracy will also be evaluated in subgroups by gender and ethnicity.

Furthermore, we evaluated the clinical application of the BITSEA Problem and Competence scale, by comparing BITSEA scores with the clinical decision of the child health professionals. We hypothesized that the clinical decision of the child health professional is associated with on the one hand higher BITSEA Problem scale scores, as high BITSEA Problem scale scores are expected to be indicative of problems, and on the other hand lower BITSEA Competence scale scores, as low BITSEA Competence scale scores are expected to be indicative of problems (i.e. delays in competencies).

METHODS

Ethics statement

Only anonymous data were used and the questionnaires were completed on a voluntary basis by the parents. Parents received written information on these questionnaires and were free to refuse to participation. Observational research with data does not fall within the ambit of the Dutch Act on research involving human subjects [14] and does not require the approval of an ethics review board. Informed consent was obtained for the use of the CBCL, since this data collection was not part of the routine health examinations. The Medical Ethics Committee of the Erasmus Medical Centre Rotterdam declared to have no objection ('formal waiver') regarding the study protocol and consent procedures.

Design and participants

The data was gathered between April 2010 and April 2011 by child health care organizations in the context of routine health examinations in the Rotterdam area, the Netherlands. Before coming to the well-child visit, parents were invited to complete the BITSEA and CBCL1.5-5. Child health professionals were trained to score the answers given by parents on the BITSEA and use the cutpoint as identified in the US [15] in their assessment whether children are at risk for, or currently experiencing, psychosocial problems. Child health professionals were blind for the answers on the CBCL. Parents of 3170 2-year-old children that attended the well-child visit handed in the questionnaire (95.5% of all parents that attended the well-child visit). Of these parents, 2184 (68.9%) parents gave informed consent for participation in the study. Children were excluded from analyses if there were too many missing items on the BITSEA Problem (>5) and Competence (>2) scales [15] (n=50) and CBCL1.5-5 (>8) (n=74), leaving a study population of 2060 (94.3%) children (mean age: 23.7 months, SD=0.7; 49.6% boys, 33.8% non-native). None of the children were under treatment of a mental health professional at the time of inclusion. Details on the design of the study are described elsewhere [11,16]. The community sample consists of 43 (2.1%) children (mean age: 23.9, SD=0.7; 51.2% boys, non-native 69.8%) that scored in the clinical range of the CBCL Total Problem score (raw score >60).

Measures

The BITSEA is designed for children aged 12 months to 36 months and consists of 42 items with three response options ('not true/rarely','somewhat true/sometimes','very true/often'). There are two multi-item scales, a Problem scale (31 items) and a Competence scale (11 items). The Problem scale assesses social-emotional/behavioural problems such as aggression, defiance, overactivity, negative emotionality, anxiety, and withdrawal. The Competence scale assesses social-emotional abilities such as empathy, prosocial behaviours, and compliance [13]. Responses should be summed for each scale, a high score on the Problem scale or a low score on the Competence scale is less favourable [15]. Previous studies showed that the BITSEA has good reliability and validity [9-11]. A study in the Netherlands yielded for the BITSEA Problem and Competence scale respectively an internal consistency (Cronbach's alpha) of 0.76 and 0.63 a test-retest reliability (intraclass correlation, ICC) of 0.75 and 0.61; parent-childcare provider interrater reliability (ICC) of 0.30 and 0.17; and significant positive correlations (Problem scale) and significant negative correlations (Competence scale) with the CBCL1.5-5 Total Problem score [11].

The CBCL1.5-5 is a well-validated [6] 100-item questionnaire designed for children aged 18 months to 5 years and has two domains (Internalising and Externalising) that are combined to give a Total Problem score. Answers are given on a 3-point scale ('not true','somewhat or sometimes true' and 'very or often true'). Children with a Total Problem score greater than 60, score in the clinical range of the CBCL1.5-5.

Ethnicity of the child was determined based on parental country of birth: a child was considered native if both parents were born in the Netherlands [17].

Clinical decision was measured as the decision of the child health professional whether or not to refer the child to a more specialized mental health professional and/or to request a follow-up consultation, as recorded in the electronic medical file. Hereinafter, we will only mention 'referral' as clinical decision, but it also entails the request for follow-up consultation.

Analyses

Mean and median BITSEA scores and CBCL1.5-5 Total Problem score

Differences in mean BITSEA Problem and BITSEA Competence scores and CBCL Total Problem score between subgroups by child gender and ethnicity are tested with independent sample t-tests. Differences in median and distribution of BITSEA Problem and BITSEA

Competence scores and CBCL Total Problem score between children with and without a clinical CBCL Total Problem score were tested with a Mann-Whitney U test for the total population and for subgroups by child gender and ethnicity. A Mann-Whitney U test was appropriate since the subgroups were small and the assumption of normality could not be met.

Screening accuracy

Screening accuracy of the BITSEA Problem scale was evaluated by calculating receiver operating characteristics (ROC) curves, with a reference group that consists of children with a CBCL Total Problem score in the clinical range. Area under the ROC curve was examined, along with, for a range of BITSEA Problem scale scores; sensitivity, specificity, positive test likelihood ratio (LHR^+) and negative test likelihood ratio (LHR^-), diagnostic odds ratio (OR) and Youden's index. All indices for screening accuracy were evaluated for the total sample as well as for boys and girls and for native and non-native children separately.

The ROC curve is a plot of sensitivity as a function of 1-specificity for all possible cutpoints of the BITSEA. The greater the area under the curve (AUC), the more discriminative power the BITSEA has in differentiating children with and without psychosocial problems. An AUC≥0.90 indicates high accuracy; 0.70≤AUC<0.90 indicates moderate accuracy; 0.50≤AUC<0.70 indicates low accuracy; and AUC=0.50 is chance level accuracy [18]. We examined the 95% confidence intervals of the AUCs to evaluate whether the screening accuracy differed significantly between subgroups.

The Youden index was calculated, which is defined as the maximum vertical distance between the ROC curve and the diagonal or chance line and is calculated as *Youden's index=sensitivity+specificity-1*. The higher the Youden index, the more optimal a cutpoint is [19].

Sensitivity is the proportion of true positives that are correctly identified by the test; specificity is the proportion of true negatives that are correctly identified by the test. To further investigate the correctness of classification, likelihood ratios were calculated. LHR^+*=sensitvitiy/(1-specificity)* is the ratio of the probability of a positive test result if the outcome is positive (true positive) to the probability of a positive test result if the outcome is negative (false positive); LHR^-*=(1-sensitivity)/specificity* is the ratio of the probability of a negative test result if the outcome is positive (false negative) to the probability of a negative test result if the outcome is negative (true negative). LHR^+>7.00 and LHR^-<0.30 indicate high screening accuracy [20].

The $OR=sensitivity*specificity/((1-sensitivity)*(1-specificity))=LHR^{+}/LHR^{-}$ of a test is the ratio of the odds of a positive test result when having the 'disorder' relative to the odds of a positive test result when not having the 'disorder'. The values of OR ranges from zero to infinity, with higher values indicating better discriminatory test performance. OR>20.00 indicate high screening accuracy [20].

The AUC, Youden's index, sensitivity, specificity, LHR^{+}, LHR^{-} and OR are independent of prevalence of the 'disorder', as opposed to the positive predictive value and negative predictive value [20].

Clinical application

The clinical application of the BITSEA was explored by evaluating the relation between BITSEA Problem and Competence scores and the decision of the child health professional to refer to another mental health professional.

Because the observations within child health professional were not independent from each other, a multilevel regression analyses was used to evaluate the relation between the clinical decision as independent variable and the BITSEA scale scores as dependent variable, corrected for confounding effects of child's gender and ethnicity. The effect sizes of the differences in mean BITSEA scale scores between children that were and were not referred were defined as $Cohen's\ d=|[mean_{worried}-mean_{not\ worried}]/SD_{worried}|$. [21] A small effect is defined as $0.20\leq d<0.50$; a medium effect is defined as $0.50\leq d<0.80$; and a large effect is defined as $d\geq 0.80$. Additionally, frequencies of referral for children scoring in the clinical range of the BITSEA were evaluated. Cutpoints are set at the 25th percentile for the Problem scale; and at 15th percentile for the Competence scale, as is specified in the manual of the BITSEA [15]

Data regarding the clinical decision of the child health professional were available for 1579 (76.7 %) children (combined data of 481 (23.3%) children were unavailable due to missing patient-codes and differences in registration methods between child health care organizations). Significant differences between the group with complete data and the group with incomplete data were found only for the mean CBCL Total Problem score ($p=0.04$). Similar differences were <u>not</u> observed for mean BITSEA Problem score, mean BITSEA Competence score, age of the child and parents, child gender and ethnicity, country of birth of the parents, person who completed the questionnaire and family composition. The effect size of the significant differences in CBCL Total Problem score between the group with complete data and the group with incomplete data, however, was small and was taken as an

indication that the data were 'missing at random' (mean CBCL Total Problem score, Cohen's d=0.10).

Multilevel regression analyses were performed in SAS software version 9.2 (SAS Institute Inc., 2009). All other analyses were performed in SPSS 19.0 (SPSS Inc. 2010).

RESULTS

Mean and median BITSEA scores and CBCL1.5-5 Total Problem score

Mean BITSEA Problem scale score, BITSEA Competence scale score and CBCL Total Problem score and standard deviations for the total population and subgroups by child gender and ethnicity are presented in Table 4.1. All mean scores differed between boys and girls and native and non-native children ($p<0.01$), except for the mean CBCL Total Problem score between boys and girls ($p=0.96$).

Table 4.2 presents the median scores and 25 and 75 percentile for children with and without a CBCL Total Problem score in the clinical range, for the total subpopulations as well as for subgroups by child gender and ethnicity. Between the subpopulations of children with and without a CBCL Total Problem score in the clinical range, all distributions and median scores differed significantly ($p<0.05$), except for the median BITSEA Competence score between girls ($p=0.18$) and native children ($p=0.22$). Within the subpopulation of children *without* a CBCL Total Problem score in the clinical range, all distributions and median scores differed between subgroups of gender and ethnicity ($p<0.01$), except for the CBCL Total Problem score distribution ($p=0.96$) and median ($p=0.87$) between boys and girls.

Table 4.1 Means and standard deviations of BITSEA Problem and Competence scores and CBCL Total Problem score for the total population and for subgroups by child gender and ethnicity

Mean (SD)	Total	Boys[1]	Girls[1]	Native	Non-native
	N=2060 (100%)	n=1021 (49.6%)	n=1033 (50.1%)	n=1364 (66.2%)	n=696 (33.8%)
CBCL Total Problem	19.1 (15.5)	19.1 (15.5)	19.1 (15.4)	16.8 (13.6)*	23.7 (17.7)*
BITSEA Problem	7.7 (5.2)	8.1 (5.4)*	7.3 (4.9)*	6.8 (4.4)*	9.4 (6.1)*
BITSEA Competence	17.8 (2.9)	17.5 (2.9)*	18.1 (2.9)*	18.2 (2.6)*	16.9 (3.3)*

[1] Percentages do not sum to 100 because of missing values.
* mean scores differed significantly between boys & girls and native & non-native children, $p<0.05$.

Table 4.2 Median and 25-75 percentile of BITSEA Problem and Competence scores and CBCL Total Problem score for children with and without a CBCL Total Problem score in the clinical range, for the total subpopulations and for subgroups by child gender and ethnicity

		CBCL Total Problem score < clinical range n=2017 (97.9%)					CBCL Total Problem sore > clinical range n=43 (2.1%)				
		Total	Boys[1]	Girls[1]	Native	Non-native	Total	Boys[1]	Girls[1]	Native	Non-native
		n=2017 (100%)	n=999 (49.5%)	n=1012 (50.2%)	n=1351 (67.0%)	n=696 (33.0%)	n=43 (100%)	n=22 (51.2%)	n=21 (48.8%)	n=13 (30.2%)	n=30 (69.8%)
CBCL Total Problem	mdn	15.0*	15.0*	15.0*	14.0*†	20.0*†	71.0*	71.5*	71.5*	75.0*	69.5*
	25%-75%	7.0-26.0*	8.0-26.0*	7.0-26.0*	6.0-23.0*†	10.0-32.0*†	66.0-76.0*	64.8-78.8*	66.5-75.0*	67.5-77.5*	64.0-75.0*
BITSEA Problem	mdn	7.0*	7.0*†	6.0*†	6.0*†	8.0*†	20.0*	19.5*	20.0*	20.0*	21.5*
	25%-75%	4.0-10.0 *	4.0-10.0*†	4.0-10.0*†	4.0-9.0*†	5.0-12.0*†	17.0-25.0*	16.8-26.5*	17.0-24.5*	18.0-21.5*	16.0-26.0*
BITSEA Competence	mdn	18.0*	18.0*†	19.0†	19.0†	17.0*†	16.0*	15.0*	16.0	16.0	15.5*
	25%-75%	16.0-20.0*	16.0-20.0*†	17.0-20.0*†	17.0-20.0*†	15.0-19.0*†	14.0-18.0*	11.0-17.0*†	15.0-19.0*†	15.5-19.0*	13.5-17.3*

[1] Percentages do not sum to 100 because of missing values.

* significant difference in distribution and median scores between subpopulations with and without a CBCL Total Problem score in the clinical range, $p<0.05$.

† significant difference in distribution and median scores within subpopulations with and without CBCL Total Problem score in the clinical range, between boys & girls and native & non-native children, $p<0.05$.

Within the subpopulation of children *with* a CBCL Total Problem score in the clinical range, all distributions and median scores did *not* differ between subgroups of gender and ethnicity ($p>0.05$), except for the Competence score distribution ($p=0.03$) between boys and girls.

Screening accuracy

In Table 4.3 are the AUC and sensitivity, specificity, LHR^+, LHR^-, OR and Youden's index presented for a range of BITSEA Problem score cutpoints, for the total population and for subgroups by gender and ethnicity.

The AUCs (95% Confidence Intervals: 95% CI) of the Problem scale was for the total population 0.97 (0.95-0.98), for boys 0.95 (0.92-0.98), for girls 0.98 (0.97-0.99), for native children 0.98 (0.97-1.00), for non-native children 0.94 (0.91-0.97). There were no significant differences in AUC between subgroups of gender and ethnicity (i.e. no (unrounded) overlapping confidence intervals). The ROC curve of the BITSEA Problem scale for the total population is presented in Figure 4.1.

The Youden index indicated the same optimal BITSEA Problem scale cutpoint for boys and girls (score 14), whereas a different optimal cutpoint was indicated by the Youden index for native children (score 17) and non-native children (score 14).

Clinical application

Of the 1579 children with complete data of both the parent and the child health professional, child health professionals referred 96 (6.1%) children for further evaluation. Both regression coefficients were significant ($p<0.05$), BITSEA Problem scale, $B=2.5$, BITSEA Competence scale, $B=-0.7$. Mean BITSEA scores differed significantly ($p<0.05$) between children that were referred (Problems scale, $M=10.1$, $SD=6.7$; Competence scale, $M=17.2$, $SD=3.2$) and children that were not referred (Problem scale, $M=7.5$, $SD=4.9$; Competence scale $M=17.9$, $SD=2.9$). The effect sizes of the differences in mean BITSEA scale scores between children that were and were not referred were for the BITSEA Problem scale $d=0.39$ and for the BITSEA Competence scale $d=0.22$. See Table 4.4.

Of the children with a score in the clinical range on the Problem scale or Competence scale or CBCL respectively 9.5%; 7.4% and 26.7% were referred.

Table 4.3 Screening accuracy for a range of BITSEA Problem scores, relative to a CBCL Total problem score in the clinical range

	Score	Sensitivity	Specificity	LHR^+	LHR^-	OR	Youden's index
Total, N=2060	AUC=0.97 (95% CI=0.95-0.98)						
	11	0.98	0.78	4.39	0.03	146.67	0.75
	12	0.98	0.82	5.58	0.03	197.98	0.80
	13	0.95	0.87	7.20	0.05	134.36	0.82
	14	**0.95**	**0.90**	**9.38**	**0.05**	**181.20**	**0.85**
	15	0.88	0.92	11.28	0.13	89.42	0.81
	16	0.84	0.94	14.94	0.17	86.65	0.78
	17	0.79	0.96	18.54	0.22	84.82	0.75
Boys, n=1021	AUC=0.95 (95% CI=0.92-0.98)						
	11	0.95	0.75	3.89	0.06	64.63	0.71
	12	0.95	0.81	5.05	0.06	90.00	0.77
	13	0.91	0.85	6.18	0.11	57.96	0.76
	14	**0.91**	**0.88**	**7.70**	**0.10**	**74.66**	**0.79**
	15	0.82	0.90	8.60	0.20	42.82	0.72
	16	0.82	0.93	12.57	0.19	64.66	0.75
	17	0.77	0.95	14.85	0.24	61.92	0.72
Girls, n=1033	AUC=0.98 (95% CI=0.97-0.99)						
	11	1.00	0.80	4.99	0.00	x	0.80
	12	1.00	0.84	6.21	0.00	x	0.84
	13	1.00	0.88	8.50	0.00	x	0.88
	14	**1.00**	**0.92**	**11.77**	**0.00**	**x**	**0.92**
	15	0.95	0.94	15.55	0.05	306.45	0.89
	16	0.86	0.95	18.46	0.15	123.19	0.81
	17	0.81	0.97	24.83	0.20	126.08	0.78
Native, n=1364	AUC=0.98 (95% CI=0.97-1.00)						
	14	0.92	0.93	13.86	0.08	168.13	0.86
	15	0.92	0.95	19.19	0.08	237.42	0.87
	16	0.92	0.97	29.00	0.08	365.02	0.89
	17	**0.92**	**0.98**	**40.23**	**0.08**	**510.97**	**0.90**
	18	0.85	0.98	42.34	0.16	269.70	0.83
	19	0.69	0.98	44.54	0.31	142.50	0.68
	20	0.54	0.99	55.96	0.47	120.08	0.53
Non-native, n=696	AUC=0.94 (95% CI=0.91-0.97)						
	11	0.97	0.67	2.93	0.05	58.79	0.64
	12	0.97	0.73	3.56	0.05	77.71	0.69
	13	0.97	0.79	4.67	0.04	110.96	0.76
	14	**0.97**	**0.83**	**5.60**	**0.04**	**138.95**	**0.79**
	15	0.87	0.86	6.21	0.15	40.05	0.73
	16	0.80	0.89	7.61	0.22	34.06	0.69
	17	0.73	0.92	8.88	0.29	30.55	0.65

Note: AUC=area under the curve; LHR^+=liklihood ratio positive test; LHR^-=liklihood ratio negative test; OR= diagnostic odds ratio. All AUCs were significant ($p<0.01$). Scores with the highest unroundend Youden's index are indicated in bold.

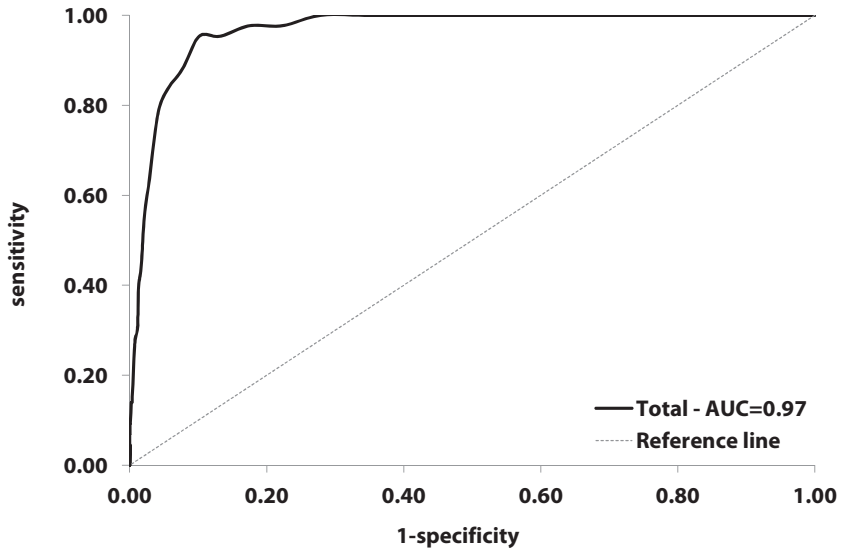


Figure 4.1 Receiver Operating Characteristic (ROC) curve for BITSEA Problem scores for the total population, relative to CBCL1.5-5 Total Problem score in the clinical range. AUC=area under the curve.

Table 4.4 Clinical application of the BITSEA; relation between BITSEA scores and the decision to refer to a specialist or request a follow-up consultation, n=1579

BITSEA scales	Referral or follow-up decision Mean (SD)		Beta[1]	Effect size[2]
	Not referred (N=1483)	Referred (N=96)		
Problem scale	7.5 (4.9)	10.1 (6.6)	2.5	0.39[a]
Competence scale	17.9 (2.9)	17.2 (3.2)	-0.7	0.22[a]

[1] Significant unstandardized Betas ($p<0.05$) are corrected for confounding effects of child's gender and ethnicity.
[2] Difference of the means divided by SD in the subgroup 'referred'.
[a] indicates a small effect ($0.2 \leq d < 0.5$).

DISCUSSION

The present study evaluated the screening accuracy of the BITSEA Problem scale for a large community sample in comparison to a subsample of children with a CBCL Total Problem score in the clinical range. Furthermore, we evaluated the clinical application of the BITSEA Problem and Competence scale. Our results indicate that the BITSEA Problem scale has high screening accuracy and that BITSEA scores were less favourable for children that were referred.

Screening accuracy

The BITSEA Problem scale has a good screening accuracy when compared to a group of children with a CBCL Total Problem score in the clinical range (i.e. AUC>90). The study performed in the US [9] found comparable sensitivity for the BITSEA Problem scale as in our study, whereas the specificity in our study was higher.

In our study the Youden index yielded the same optimal cutpoint for boys and girls (score 14). In the US-study, for the age range 24-29 months, score 14 was also identified as the cutpoint for boys, whereas the cutpoint for girls was 13. These different results between studies might be attributed to different characteristics of the study populations and to the different methods of indicating a cutpoint. Also, as opposed to the US-study, in our study completion of the BITSEA was not anonymous, since the answers were used by the child health professional to assess the child's development.

In our study we found different optimal cutpoints for native and non-native children, where native children differed from the other (sub)samples in cutpoint as indicated by the Youden index; score 17. The mean BITSEA Problem scores differed significantly between boys and girls and native and non-native children, but the difference in mean BITSEA Problem scores between native and non-native children was larger (effect size=$(mean_{non\text{-}native}-mean_{native})/sd_{non\text{-}native}$= 0.43) than the difference in mean BITSEA Problem scores between boys and girls (effect size=$(mean_{boys}-mean_{girls})/sd_{boys}$=0.15), which might explain the different optimal cutpoints between native and non-native children and not between boys and girls. The outcome that the screening accuracy of the BITSEA is the same for native and non-native children is valuable. However, the application of different cutpoints for different ethnic groups in preventive child health care may not be desirable, since it is difficult to determine whether the different distribution and mean BITSEA scores can be attributed to the actual amount or seriousness of problems, or that it reflects cultural differences (e.g. in interpretation of behaviour, or question items). Moreover, the composition of ethnic groups may change over time, which would mean the (continuous) evaluation and adjustment of cutpoints.

The BITSEA Competence scale was excluded from screening accuracy analyses because the content of the items of the BITSEA Competence scale do not resemble the content of the items on the CBCL Total Problem score. This is supported by the low and non-existence of correlations between the mean BITSEA Competence scores and the CBCL Total Problem score found in prior studies [9-11]. The decision not to include the BITSEA Competence scale in the analyses seems also (partly) justified by the results of the present study that the

median BITSEA Competence score did not differ between girls and native children with and without a CBCL Total Problem score in the clinical range.

Clinical application

The BITSEA Problem and Competence score were significant, respectively positively and negatively, associated with the child health professional's decision whether or not referral was required. These results indicate that scores were less favourable for children who were referred, compared to children who were not referred. However, the difference in mean BITSEA Problem and Competence scores were small ($0.20 \leq d < 0.50$).

The child health professionals referred 7.4-9.5% of the children with a score in the clinical range on either BITSEA scale and 26.7% of the children that score in the clinical range of the CBCL1.5-5. The frequency of children that were referred was relatively small (n=96). Moreover, only 30 (2%) children of whom we had complete data of the parent and child health professional, had a score in the clinical range of the CBCL1.5-5 Total Problem score, of which 8 were referred. These small frequencies might have caused a power problem. Other studies found percentages, comparable to our referral frequencies on the CBCL1.5-5. In one study child health professionals referred (or requested a follow-up consultation) 22.4% children with a high score on both the parent and teacher completed SDQ (>P90) [22]. In two other studies child health professionals referred (or requested a follow-up consultation) 19% of the children with a score in the clinical range on the CBCL [23] and ITSEA [24]. However, in these latter two studies the child health professional was blind to the questionnaire score, as were the professionals in our study for the CBCL1.5-5, this might also partly explain the difference in frequencies. Not all children with a score in the clinical range on an early detection instrument are referred, possibly because the problematic emotions or behaviours are mild or are considered to be temporarily (e.g. after a major life event). Then, the child health professional may offer advice about how to cope with the circumstances instead of referring the child to more specialized care [15]. Also, the degree of concerns the parents have about their child's development is likely to play a role in the clinical decision of the child health professionals, since child health providers are found to be more likely to refer when parents are concerned about their child's behaviour [25,26].

Strengths and limitations

A major strength of our study is that the analyses of screening accuracy were performed on a large and diverse community sample, which adds to the power of the study. Additionally, the answers on the BITSEA were not anonymous, since the child health professional used the BITSEA to assess the child's development during the well-child visit, this could be seen as either a strength or a limitation. The parents could have completed the BITSEA more seriously; on the other hand it could also have led to more socially desirable answers.

Our study also has some limitations. First, in our sample a low percentage (i.e. 2.1%) of children had a CBCL Total Problem score in the clinical range, whereas based on the literature a higher percentage (i.e. 6.5-12.5%) was expected [23,24,27]. This might be indicative of a response bias: not all parents with children with (serious) psychosocial problems may come to the well-child visit, possibly because they are already under treatment of a specialized mental health professional, or because they did not wish to participate in the study. Different cutpoints might be the result of the response bias, as opposed to when the sample consisted of more children with CBCL Total Problem scores in the clinical range. However, the percentage of parents that attended the well-child visit and also completed the questionnaire is quite high (i.e. 95.5%), indicating that the sample is a good reflection of the population in the Rotterdam area that make use of the preventive child health care, may complete the questionnaire in the future and on whom the cutpoints should be applied. However, as a consequence, the subgroups of child gender and ethnicity in the 'clinical range sample' are quite small. This does not lead to large confidence intervals since the confidence intervals are calculated based on the large total study population.

Another limitation of our study is the use of children with a CBCL Total Problem score in the clinical range, a subsample of the community sample, as a reference group for the ROC analyses. This excluded the possibility to evaluate the screening accuracy of the BITSEA Competence scale. Moreover, the criterion-related validity of the CBCL (criterion in this case being the presence of psychosocial problems) might limit the quality of findings on screening accuracy. However, the CBCL1.5-5 is a well validated questionnaire and often used as a gold standard for research and clinical work among broad-band early detection instruments for psychosocial problems.

The study was performed in the Netherlands with a Dutch population; this might hamper generalizations to populations of other cultures. However, no difference in screening accuracy was found between native and non-native Dutch children. Moreover, a previous study also

showed no difference between native and non-native Dutch children in reliability and validity of the BITSEA [11], suggesting that the BITSEA performs equally well for samples of different cultures.

Future research

We recommend future studies to evaluate the screening accuracy of both the BITSEA Problem and Competence scale with a reference group of children with a broad range of psychosocial problems who are diagnosed by a mental health professional. Additionally, evaluating the clinical application of the BITSEA in a larger sample, and including the concerns of parents regarding their child's development in these analyses, might provide more insight in the value of the BITSEA in the preventive child health care. Furthermore, we recommend evaluating the application of the determined BITSEA cutpoints by child health professionals and the subsequent adherence of the referrals by parents.

CONCLUSION

The BITSEA Problem scale shows a good screening accuracy with regard to psychosocial problems as indicated by the CBCL1.5-5, for the total population and for subgroups of child gender and ethnicity. Furthermore, the clinical application of the BITSEA was as hypothesized; less favourable scores for children that were referred, compared to children that were not referred. These results indicate that the BITSEA may be suitable for use in the preventive child health care.

REFERENCES

1. Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP (2001) Identification and management of psychosocial problems by preventive child health care. Archives of Pediatrics & Adolescent Medicine 155: 462-469.
2. Laurent de Angulo MS, Brouwers-de Jong EA, Bulk A (2005) Ontwikkelingsonderzoek in de jeugdgezondheidszorg. Assen: Koninklijke Van Gorcum.
3. Sturner RA (1991) Parent questionnaires - Basic office equipment. Journal of Developmental and Behavioral Pediatrics 12: 51-54.
4. Glascoe FP (2005) Screening for developmental and behavioral problems. Mental Retardation and Developmental Disabilities Research Reviews 11: 173-179.
5. Jellinek MS, Murphy JM (1999) Screening for psychosocial disorders in pediatric practice. American Journal of Diseases of Children 142: 1153-1157.
6. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington:VT: University of Vermont, Research Center for Children, Youth, and Families.
7. Briggs-Gowan MJ, Carter AS (1998) Preliminary acceptability and psychometrics of the Infant-Toddler Social and Emotional Assessment (ITSEA): A new adult-report questionnaire. Infant Mental Health Journal 19: 422-445.
8. Carter AS, Briggs-Gowan MJ, Jones SM, Little TD (2003) The Infant-Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. Journal of Abnormal Child Psychology 31: 495-514.
9. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.
10. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, et al. (2009) The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behavior & Development 32: 291-297.
11. Kruizinga I, Jansen W, de Haan CL, van der Ende J, Carter AS, et al. (2012) Reliability and validity of the dutch version of the brief infant-toddler social and emotional assessment (BITSEA). PLoS ONE 7.
12. Karabekiroglu K, Briggs-Gowan MJ, Carter AS, Rodopman-Arman A, Akbas S (2010) The clinical validity and reliability of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). Infant Behavior & Development 33: 503-509.
13. Briggs-Gowan MJ, Carter AS (2008) Social-emotional screening status in early childhood predicts elementary school outcomes. Pediatrics 121: 957-962.
14. Wet medisch-wetenschappelijk onderzoek met mensen [Dutch Act on research involving human subjects] (1998). Bwb-id:BWBR0009408. http://wetten.overheid.nl/BWBR0009408/geldigheidsdatum_05-02-2010.

15. Briggs-Gowan MJ, Carter AS (2002) Brief Infant-Toddler Social and Emotional Assessment (BITSEA) mannual, version 2.0. New Haven: CT:Yale University.

16. Kruizinga I, Jansen W, Carter AS, Raat H (2011) Evaluation of an early detection tool for social-emotional and behavioral problems in toddlers: The Brief Infant Toddler Social and Emotional Assessment - A cluster randomized trial. Bmc Public Health 11.

17. Definitions 'native' and 'immigrant' (2012) - http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?conceptid=37. [Statistics Netherlands].

18. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

19. Akobeng AK (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatrica 96: 644-647.

20. Fischer JE, Bachmann LM, Jaeschke R (2003) A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Medicine 29: 1043-1051.

21. Cohen J (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erblaum Associates.

22. Bevaart F (2013) Ethnic Differences in Pathways in Care for Young Children with Problem Behaviour: Road Work in Progress.

23. Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP (2004) Identification and management of psychosocial problems among toddlers in Dutch preventive child health care. Archives of Pediatrics & Adolescent Medicine 158: 811-817.

24. Klein Velderman M, Crone MR, Wiefferink CH, Reijneveld SA (2010) Identification and management of psychosocial problems among toddlers by preventive child health care professionals. European Journal of Public Health 20: 332-338.

25. Ellingson KD, Briggs-Gowan MJ, Carter AS, Horwitz SM (2004) Parent identification of early emerging child behavior problems - Predictors of sharing parental concern with health providers. Archives of Pediatrics and Adolescent Medicine 158: 766-772.

26. Glascoe FP, Maclean WE, Stone WL (1991) The importance of parents concerns about their childs behavior. Clinical Pediatrics 30: 8-11.

27. Koot HM,Verhulst FC (1991) Prevalence of problem behavior in Dutch children aged 2-3. Acta Psychiatrica Scandinavica 83: 1-37.

Chapter 5

Screening for autism spectrum disorders with the brief infant-toddler social and emotional assessment

Ingrid Kruizinga
Janne C. Visser
Tamara van Batenburg-Eddes
Alice S. Carter
Wilma Jansen
Hein Raat

Published as:
Kruizinga I, Visser JC, van Batenburg-Eddes T, Carter AS, Jansen W, et al. (2014) Screening for Autism Spectrum Disorders with the Brief Infant-Toddler Social and Emotional Assessment. PLoS ONE 9(5): e97630. doi:10.1371/journal.pone.0097630

ABSTRACT

Background: Using parent-completed questionnaires in (preventive) child health care can facilitate the early detection of psychosocial problems and psychopathology, including autism spectrum disorders (ASD). A promising questionnaire for this purpose is the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). The screening accuracy with regard to ASD of the BITSEA Problem and Competence scales and a newly calculated Autism score were evaluated.

Methods: Data, that was collected between April 2010 and April 2011, from a community sample of 2-year-olds (N=3127), was combined with a sample of preschool children diagnosed with ASD (N=159). For the total population and for subgroups by child's gender, area under the Receiver Operating Characteristic (ROC) curve was examined, and across a range of BITSEA Problem, Competence and Autism scores, sensitivity, specificity, positive and negative likelihood ratio's, diagnostic odds ratio and Youden's index were reported.

Results: The area under the ROC curve (95% confidence interval, [95%CI]) of the Problem scale was 0.90 (0.87-0.92), of the Competence scale 0.93 (0.91-0.95), and of the Autism score 0.95 (0.93-0.97). For the total population, the screening accuracy of the Autism score was significantly better, compared to the Problem scale. The screening accuracy of the Competence scale was significantly better for girls (AUC=0.97; 95%CI=0.95-0.98) than for boys (AUC=0.91; 95%CI=0.88-0.94).

Conclusion: The results indicate that the BITSEA scales and newly calculated Autism score have good discriminative power to differentiate children with and without ASD. Therefore, the BITSEA may be helpful in the early detection of ASD, which could have beneficial effects on the child's development.

INTRODUCTION

Preventive child health care offers a systematic opportunity for the early detection of psychosocial problems and psychopathology, such as autism spectrum disorders (ASD), among toddlers. In the Netherlands, preventive child health care for children of ages 0-4 years is delivered through community well-child clinics that provide routine developmental assessment and vaccinations (i.e. well-child visits) and that are free of charge [1].

ASD represents a set of neurodevelopmental disorders that are characterized by impairments in the domains of reciprocal social interactions and communication and by restrictive, stereotyped patterns of behaviour [2]. In the current Diagnostic and Statistical Manual of Mental disorders, 5th edition, ASD's are part of the pervasive developmental disorders and classified into three main categories, namely: autistic disorder, Asperger's disorder and pervasive developmental disorder-not otherwise specified [2]. Studies report ASD prevalence rates of about 1.0% [3,4]. Abnormal functioning that is indicative of ASD starts before 3 years of age [2]. On average, the first symptoms to arouse parental concerns about children eventually diagnosed with ASD occur before the second birthday. However, the average age of ASD diagnosis is approximately three years of age and often occurs later [5]. These findings suggest that it should be possible to detect and diagnose ASD earlier. Early detection of ASD is important because early access to interventions may improve children's outcomes [6,7], and diagnosis may enhance parent's understanding and coping with the impairments of their child [8].

One approach for facilitating early identification of ASD is the population-based screening of children as part of well-child visits using parent-completed questionnaires [9,10] Several instruments are developed for the early detection of ASD, of which the use of the Checklist for Autism in Toddlers (CHAT) [11] and the Modified Checklist for Autism in Toddlers (M-CHAT) [12] is advocated by autism support organizations [13]. However, early detection instruments that are used in a preventive health care setting should cover a broad range of psychosocial problems, since limited time and capacity in the preventive child health care make it undesirable to screen for each psychosocial problem separately. Also, it has been shown that psychosocial problems tend to co-occur [14,15], and that individual problems may apply to more than one disorder [16]. In addition to measuring problem domains, it is crucial to also measure competence domains. Delays in the acquisition of competencies are strongly related to a wide range of psychosocial problems later in life [17] and are often the prodromal signs of developmental disorders, such as ASD [18].

The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) [19] is a promising and short (42 items) questionnaire, that measures both problems (Problem scale) and delays in the acquisition of competencies (Competence scale) in 1-3 year olds, and also consists of items designed to measure ASD symptoms. The BITSEA is not designed to diagnose ASD, but it may be useful as a screener for identifying children with this disorder [20]. Previous studies have shown that the BITSEA Problem and Competence scale has adequate reliability for the Problem scale and validity for the Problem and Competence scale [19,21-23]. The study performed in the Netherlands [23] evaluated among others the internal consistency, test-retest reliability, concurrent validity and discriminant validity. An adequate Cronbach's alpha (i.e. >0.70 [24]) was found for the Problem scale (0.76) and marginal for the Competence scale (0.63). Test-retest reliability was adequate (>0.70 [25]) for the Problem scale (0.75) and marginal for the Competence scale (0.61). The BITSEA Problem scale was positively correlated with the CBCL, Pearson coefficients of 0.66 (Internalizing), 0.65 (Externalizing) and 0.75 (Total Problem). The BITSEA Competence score was negatively correlated with the CBCL, Pearson coefficients of -0.26 (Internalizing), -0.23 (Externalizing) and -0.26 (Total Problem). All correlations were significant ($p<0.01$). The mean BITSEA score was compared between a group of parents that worried about the development of their child and a group that did not worry. The Problem and Competence score were significantly less favourable in the group of parents that worried, compared to the group of parents that did not worry (effect sizes were respectively 0.93 and 0.52).

Also the sensitivity and specificity of the BITSEA has been evaluated in several studies [19,26,27] One study, conducted in the United States [19], examined its sensitivity and specificity in a community sample of 1280 children. In this study, children with scores in the clinical range on the Child Behavioral Checklist (CBCL1.5-5) [28] and Infant-Toddler Social and Emotional Assessment (ITSEA) [29,30] were used as reference groups for the evaluation of the Problem scale. A sensitivity of respectively 93.2% and 78.1% and a specificity of respectively 78.0% and 88.8% were found. The Competence scale was evaluated against a group of children with a score in the clinical range on the ITSEA and had a sensitivity of 68.9% and a specificity of 95.1%. Problem scale cutpoints were chosen at scores of $\geq 75^{th}$ percentile and Competence scale cutpoints were chosen at scores of $<15^{th}$ percentile [31]. In a Turkish study [26], in a community sample of 462 children, sensitivity and specificity of only the Competence scale was examined relative to children treated in a child psychiatry outpatient clinic with an autism diagnosis (n=35). In this study, the sensitivity was 72%-93% and specificity was 76%-85%, depending on the cutpoint chosen. A Dutch study [27] evaluated

the screening accuracy of the BITSEA Problem scale more extensively than prior studies. The screening accuracy was evaluated with multiple indices (i.e. area under the curve, sensitivity, specificity, likelihood ratio's, diagnostic odds ratios and Youden's index) by calculating receiver operating characteristic (ROC) curves of the BITSEA Problem scale relative to the CBCL Total Problem scale. Indices of screening accuracy for a range of BITSEA Problem scores were presented, because different cutpoints might be chosen in different settings (e.g. clinical application versus epidemiological research). In that study, the screening accuracy of the BITSEA Competence scale was not evaluated with a reference group of children with a CBCL Total Problem score in the clinical range, since the CBCL Total Problem score does not measure competencies.

In the present study we aim to evaluate the screening accuracy of both the BITSEA Problem and Competence scales with regard to an ASD diagnosis. Additionally, we will evaluate the screening accuracy of the BITSEA items that are specifically intended to signal ASD, since little is known about the performance of these items in the detection of ASD. Previous studies showed differences in mean BITSEA scores between boys and girls (with boys scoring less favourably) [19,22,23], therefore the screening accuracy is also evaluated in subgroups by child gender.

METHODS

Ethics statement

Regarding the data collection of the community sample; only anonymous data were used and the questionnaires were completed on a voluntary basis by the parents. Parents received written information on these questionnaires and were free to refuse to participation. Observational research with data does not fall within the ambit of the Dutch Act on research involving human subjects [32] and does not require the approval of an ethics review board. The Medical Ethics Committee of the Erasmus Medical Centre Rotterdam declared to have no objection ('formal waiver') regarding the study protocol and consent procedures. The Medical Ethical Committee of the University Medical Centre St. Radboud Nijmegen approved the study protocol regarding the ASD-study. We are prepared to make the data available upon request.

Design and participants

For the present study, data from two separate samples were combined. First, data from a community sample of 2-year old children was used. These data were gathered between April 2010 and April 2011 by child health care organizations in the context of routine health examinations in the Rotterdam area, the Netherlands. Parents of 3170 children that attended the well-child visit handed in the questionnaire (95.5% of all parents that attended the well-child visit). Children were excluded from the analyses if there were too many missing items on both BITSEA scales [20] (n=43), leaving a study population of 3127 (94.2%) children. No children in the community sample were under treatment of a mental health professional at the time of inclusion. Details on the design and participants of the community sample are described elsewhere [23].

Second, data from a sample of children diagnosed with ASD were used (i.e. ASD-sample). Children between the ages of 12-40 months were recruited in the DIANE-study (Diagnosis and Intervention of Autism in the Netherlands) [33] at Karakter Child and Adolescent Psychiatry University Center Nijmegen, the Netherlands. Children with a positive score on the Early Screening of Autistic Traits Questionnaire [34] and/or for whom there were major concerns regarding social and communicative development entered the study between spring 2004 and spring 2007. Parents of the ASD-sample completed the ITSEA (i.e. a more comprehensive measure that includes the BITSEA items) at home before their first visit for diagnostic assessments and all children underwent an extensive psychiatric assessment (i.e. administration of the Autism Diagnostic Observation Schedule and Autism Diagnostic Interview-Revised) observations of standardised parent-child play and standardised assessment of cognitive and language skills). Details on the design and participants of the ASD-sample are described elsewhere [35]. For the purpose of this study, answers on BITSEA items were extracted from the larger pool of ITSEA items. Children were excluded from the analyses if they did not receive a diagnosis (n=29), if they received a diagnosis other than ASD (n=69) (i.e. false positives), if there were too many missing items on the BITSEA scales [20] (n=19), or if they were younger than 12 months (n=2) leaving a study population of 159 (57%) children.

Measures

The BITSEA, designed for 1-to-3-year old children, consists of 42 items with three response options ('not true/rarely' (0), 'somewhat true/sometimes' (1), 'very true/often' (2)) and

comprises two multi-item scales; a Problem scale (31 items) and a Competence scale (11 items). The Problem scale assesses social-emotional/behavioural problems such as aggression, defiance, overactivity, negative emotionality, anxiety, and withdrawal. The Competence scale assesses social-emotional abilities such as empathy, prosocial behaviours, and compliance [31]. Responses can be summed for each scale: a high score on the Problem scale and/or a low score on the Competence scale is less favourable [20]. The BITSEA also consists of 17 items that are specifically included for the early detection of ASD belonging to either the Problem scale (9 items) or the Competence scale (8 items). The autism items reflect problems behaviours that are typical of children with ASD (e.g. *put things in a special order over and over*) and competencies in which deficits are often present in children with ASD (e.g. *points to show you something far away*) [20]. Although these items formally do not represent a separate scale, we calculated the Autism score analogous to the Problem scale score, yielding a good internal consistency (Cronbach's alpha=0.77). Answers on the autism items belonging to the Competence scale were first reversed before all autism items were summed, so a higher Autism score would represent more problems and fewer competencies. Children with more than 3 missing items were excluded for analyses (n=48). Excluded children were all part of the community sample.

Items on standard socio-demographic variables were included: child age and gender.

Analyses

Demographic characteristics and mean BITSEA scores

Differences in mean BITSEA scores and child age between the community sample and the ASD-sample were tested with independent sample t-tests. Differences in gender composition of the community sample and ASD-sample were tested with Chi-square tests.

Screening accuracy

Screening accuracy was evaluated by calculating receiver operating characteristic (ROC) curves, with a reference group that consists of children with a diagnosis of ASD. The area under the ROC curve was examined, along with, for a range of Problem and Competence scale scores and the Autism score; sensitivity, specificity, positive test likelihood ratio (LHR^+) and negative test likelihood ratio (LHR^-), diagnostic odds ratio (OR) and Youden's index. All indices for screening accuracy were evaluated for the total sample as well as for boys and girls separately.

The ROC curve is a plot of sensitivity as a function of 1-specificity for all possible cutpoints of the BITSEA. The greater the area under the curve (AUC), the more discriminative power the BITSEA has in differentiating children with and without ASD. An AUC>0.90 indicates high accuracy; 0.70≤AUC<0.90 indicates moderate accuracy; 0.50≤AUC<0.70 indicates low accuracy; and AUC=0.50 is chance level accuracy [36]. We examined the 95% confidence intervals of the AUCs to evaluate whether the screening accuracy differed significantly between subgroups.

To determine the optimal cutpoint, the Youden index was used, which is defined as the maximum vertical distance between the ROC curve and the diagonal or chance line and is calculated as *Youden's index=sensitivity+specificity-1* [37].

Sensitivity is the proportion of true positives that are correctly identified by the test; specificity is the proportion of true negatives that are correctly identified by the test. To further investigate the correctness of classification, likelihood ratios were calculated. LHR^{+}*=sensitvitiy/(1-specificity)* is the ratio of the probability of a positive test result if the outcome is positive (true positive) to the probability of a positive test result if the outcome is negative (false positive); LHR^{-}*=(1-sensitivity)/specificity* is the ratio of the probability of a negative test result if the outcome is positive (false negative) to the probability of a negative test result if the outcome is negative (true negative). LHR^{+}>7.00 and LHR^{-}<0.30 indicate high screening accuracy [38].

The *OR=sensitivity*specificity/((1-sensitivity)*(1-specificity))=*LHR^{+}/LHR^{-} of a test is the ratio of the odds of a positive test result when having the 'disorder' relative to the odds of a positive test result when not having the 'disorder'. The values of OR ranges from zero to infinity, with higher values indicating better discriminatory test performance. OR>20.00 indicate high screening accuracy [38].

The AUC, Youden's index, sensitivity, specificity, LHR^{+}, LHR^{-} and OR are independent of prevalence of the 'disorder', as opposed to the positive predictive value and negative predictive value, therefore the latter were not evaluated in this study [38].

All analyses were performed in SPSS 20.0 (SPSS Inc. 2011).

RESULTS

The demographic characteristics of the multiethnic community sample and ASD-sample are presented in Table 5.1. In comparison to the community sample, the ASD-sample consisted of older children ($t=58.3$, $p<0.001$) and more boys ($X^2=50.2$, $p<0.001$).

Mean BITSEA scores

The mean Problem and Competence scale scores and the Autism score are presented in Table 5.1. In comparison to children in the community sample, children in the ASD-sample scored less favourably on the Problem scale ($t=28.1$, $p<0.001$), the Competence scale ($t=29.9$, $p<0.001$) and Autism score ($t=37.3$, $p<0.001$).

Table 5.1 Child characteristics of the autism spectrum disorder (ASD) sample and community sample

	ASD-sample N=159	Community sample N=3127
Gender[a]*		
Boys, % (N)	79.2 (126)	50.0 (1564)
Girls, % (N)	20.8 (33)	49.1 (1535)
Age (months)*, M (SD)	31.8 (6.4)	23.7 (0.7)
BITSEA Problem scale score*, M (SD)	20.5 (8.7)	7.8 (5.3)
BITSEA Competence scale score*, M (SD)	10.0 (4.0)	17.5 (3.0)
BITSEA Autism score*, M (SD)	14.6 (5.2)	4.1 (3.3)

a. Percentages do not sum to 100% due to missing values.
* Significant differences in composition between ASD-sample and community sample with regard to gender, and age and mean Problem scale score, Competence scale score, and Autism score, p<0.001.

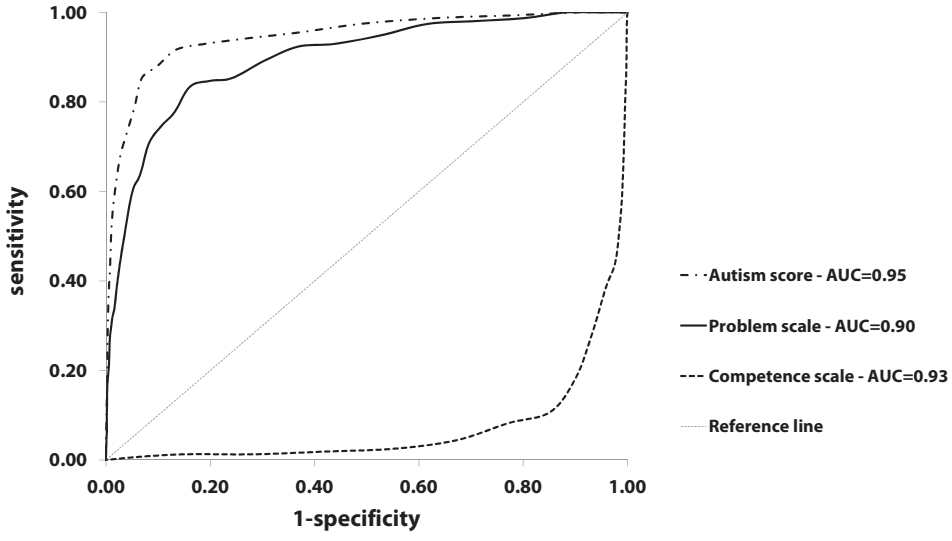


Figure 5.1 ROC curves and AUC of the BITSEA Problem and Competence scale and BITSEA Autism score relative to a sample of children with a diagnosis of autism spectrum disorder.

Screening accuracy

ROC curves of the Problem and Competence scale scores and Autism score are presented in Figure 5.1. In Table 5.2, the AUC and sensitivity, specificity, LHR^+, LHR^-, OR and Youden's

Table 5.2 The screening accuracy of the BITSEA scales with regard to autism spectrum disorders: Area Under the Curve and sensitivity, specificity, likelihood ratios, diagnostic odd ratio and Youden's index for a range of Problem and Competence scores, for the total sample and for subgroups by gender

Scale	BITSEA Problem							BITSEA Competence						
	Score	Sens	Spec	LHR^+	LHR^-	OR	J	Score	Sens	Spec	LHR^+	LHR^-	OR	J
Total	AUC=0.90 (95%CI=0.87-0.92)							AUC=0.93 (95%CI=0.91-0.95)						
N=3286	9	0.92	0.63	2.51	0.12	20.73	0.56	11	0.98	0.56	2.20	0.04	52.97	0.53
	10	0.89	0.70	2.94	0.16	18.92	0.59	12	0.96	0.61	2.48	0.07	37.00	0.57
	11	0.85	0.76	3.53	0.19	18.24	0.61	13	0.93	0.72	3.34	0.10	35.01	0.65
	12	0.85	0.80	4.22	0.19	22.05	0.65	14	0.90	0.82	4.91	0.12	40.41	0.72
	13	**0.83**	**0.84**	**5.18**	**0.20**	**26.22**	**0.67**	**15**	**0.85**	**0.89**	**7.92**	**0.17**	**47.95**	**0.74**
	14	0.78	0.87	5.92	0.26	23.06	0.65	16	0.77	0.92	9.38	0.25	37.71	0.69
	15	0.75	0.89	7.08	0.28	24.85	0.64	17	0.67	0.96	15.19	0.34	44.37	0.63
	16	0.71	0.92	8.60	0.32	26.93	0.62	18	0.56	0.97	21.95	0.46	48.15	0.53
	17	0.64	0.93	9.75	0.39	25.10	0.57	19	0.43	0.98	22.48	0.58	38.49	0.41
Boys	AUC=0.88 (95%CI=0.85-0.91)							AUC=0.91 (95%CI=0.88-0.94)*						
N=1690	9	0.92	0.60	2.31	0.13	17.48	0.52	11	0.97	0.53	2.08	0.05	42.04	0.51
	10	0.88	0.67	2.64	0.18	14.80	0.55	12	0.95	0.60	2.34	0.09	27.19	0.54
	11	0.85	0.73	3.19	0.21	15.50	0.58	13	0.92	0.71	3.12	0.12	26.66	0.62
	12	0.84	0.78	3.88	0.20	19.17	0.62	14	0.88	0.82	4.84	0.14	33.88	0.70
	13	**0.83**	**0.82**	**4.61**	**0.20**	**22.63**	**0.65**	**15**	**0.82**	**0.88**	**6.92**	**0.20**	**34.70**	**0.71**
	14	0.77	0.85	5.18	0.27	19.18	0.62	16	0.73	0.90	7.65	0.30	25.51	0.63
	15	0.74	0.88	6.02	0.30	20.17	0.62	17	0.62	0.94	11.23	0.40	28.23	0.57
	16	0.70	0.90	7.28	0.33	21.81	0.60	18	0.50	0.97	15.66	0.52	30.15	0.47
	17	0.63	0.92	8.06	0.40	19.92	0.55	19	0.37	0.98	15.54	0.65	24.09	0.35
Girls	AUC=0.93 (95%CI=0.89-0.97)							AUC=0.97 (95%CI=0.95-0.98)*						
N=1568	9	0.94	0.66	2.79	0.10	28.74	0.60	11	0.98	0.66	2.85	0.03	91.70	0.64
	10	0.94	0.73	3.46	0.09	39.10	0.66	12	0.97	0.69	3.11	0.04	72.11	0.66
	11	0.87	0.79	4.05	0.16	24.65	0.66	13	0.95	0.78	4.32	0.07	62.63	0.73
	12	0.87	0.82	4.79	0.16	30.36	0.69	14	0.92	0.81	4.90	0.10	49.22	0.73
	13	**0.84**	**0.86**	**6.04**	**0.19**	**32.24**	**0.70**	**15**	**0.88**	**0.94**	**14.14**	**0.12**	**113.81**	**0.82**
	14	0.81	0.89	7.22	0.22	33.11	0.69	16	0.82	0.97	26.15	0.19	138.50	0.79
	15	0.77	0.91	8.94	0.25	36.14	0.69	17	0.73	1.00	x	0.27	x	0.73
	16	0.74	0.93	10.89	0.28	39.33	0.67	18	0.62	1.00	x	0.38	x	0.62
	17	0.68	0.95	12.97	0.34	38.09	0.63	19	0.49	1.00	x	0.51	x	0.49

* The Competence scale AUCs differ significantly between boys and girls (i.e. the 95% confidence intervals do not overlap).

Note: AUC=area under the curve; 95%CI=95% confidence interval; sens=sensitivity; spec=specificity; LHR^+= likelihood ratio positive test; LHR^-=likelihood ratio negative test; OR=diagnostic odds ratio; J=Youden's index. All AUCs were significant ($p<0.001$). Scores with the highest unrounded Youden's index are indicated in bold.

index are presented for a range of BITSEA scale, for the total population and for subgroups by child gender.

The AUCs (95% confidence interval [CI]) of the Problem scale was 0.90 (0.87-0.92), and of the Competence scale 0.93 (0.91-0.95). The screening accuracy of the Problem scale was

Table 5.3 The screening accuracy of the BITSEA Autism score: Area Under the Curve and sensitivity, specificity, likelihood ratios, diagnostic odds ratio and Youden's index for a range of Autism scores, for the total sample and for subgroups by gender

	BITSEA Autism score						
	Score	Sens	Spec	LHR^+	LHR^-	OR	J
Total			AUC=0.95 (95%CI=0.93-0.97)				
N=3236	6	0.94	0.72	3.43	0.08	43.33	0.67
	7	0.93	0.81	4.86	0.09	56.11	0.74
	8	0.92	0.86	6.77	0.10	70.71	0.78
	9	0.88	0.90	9.05	0.13	67.53	0.78
	10	**0.85**	**0.93**	**12.40**	**0.16**	**78.79**	**0.78**
	11	0.79	0.95	14.39	0.22	64.70	0.73
	12	0.72	0.96	19.44	0.29	66.80	0.68
	13	0.68	0.97	25.35	0.33	75.96	0.65
	14	0.59	0.98	37.59	0.42	89.38	0.57
Boys			AUC=0.93 (95%CI=0.91-0.96)				
N=1671	5	0.94	0.59	2.29	0.09	24.30	0.53
	6	0.94	0.70	3.08	0.09	33.74	0.63
	7	0.92	0.78	4.10	0.10	40.05	0.70
	8	0.90	0.84	5.66	0.11	49.92	0.74
	9	**0.88**	**0.89**	**7.73**	**0.13**	**57.56**	**0.77**
	10	0.85	0.91	9.94	0.16	60.28	0.76
	11	0.79	0.93	11.79	0.23	51.33	0.72
	12	0.70	0.95	14.20	0.32	44.76	0.65
	13	0.65	0.97	18.97	0.36	52.46	0.62
Girls			AUC=0.97 (95%CI=0.95-0.99)				
N=1543	4	1.00	0.57	2.33	0.00	x	0.57
	5	1.00	0.67	3.07	0.00	x	0.67
	6	0.97	0.76	4.00	0.04	93.93	0.73
	7	0.97	0.84	6.23	0.04	163.02	0.81
	8	**0.97**	**0.89**	**8.76**	**0.04**	**241.62**	**0.86**
	9	0.87	0.92	10.79	0.14	76.91	0.79
	10	0.87	0.95	16.46	0.14	120.83	0.82
	11	0.81	0.96	18.48	0.20	91.29	0.76
	12	0.81	0.97	32.09	0.20	161.62	0.78

Note: AUC=area under the curve; 95%CI=95% confidence interval; sens=sensitivity; spec=specificity; LHR^+=likelihood ratio positive test; LHR^-=likelihood ratio negative test; OR=diagnostic odds ratio; J=Youden's index. All AUCs were significant ($p<0.001$). Scores with the highest unrounded Youden's index are indicated in bold.

equal for girls (AUC=0.93; 95%CI=0.89-0.97) and boys (AUC=0.88; 95%CI=0.85-0.91). The screening accuracy of the Competence scale was better for girls (AUC=0.97; 95%CI=0.95-0.98) than for boys (AUC=0.91; 95%CI=0.88-0.94). The Youden index indicated the same optimal cutpoint for the total population and for boys and girls for the Problem scale (score 13) and for the Competence scale (score 15).

In Table 5.3 AUCs and sensitivity, specificity, LHR^+, LHR^-, OR and Youden's index are presented for a range of Autism scores for the total population and for subgroups by child gender. The AUC was 0.95 (0.93-0.97) and the screening accuracy was equal for girls (AUC=0.97; 95%CI=0.95-0.99) and boys (AUC=0.93; 95%CI=0.91-0.96). The Youden index indicated different optimal cutpoint for the total population (score 10) and for boys (score 9) and girls (score 8).

The scores in the general population with the highest Youden index as cutpoints for the Problem and Competence scale and Autism score yielded concern level of ASD of respectively 16.1%, 10.1% and 6.9% children.

DISCUSSION

The present study evaluated the screening accuracy of the Problem and Competence scales and the newly calculated Autism score for a community sample in comparison to a sample that consists of children with an ASD diagnosis. Our results indicate that the Problem and Competence scales and the Autism score have high screening accuracy to detect ASD (i.e. AUC>0.90).

In our study we present the sensitivity and specificity for a range of BITSEA scores, because different cutpoints might be chosen in different settings (e.g. clinical application versus epidemiological research). For the comparison of the sensitivity and specificity with results of other studies we chose to discuss the sensitivity and specificity for the optimal cutpoint as indicated by the Youden index. In comparison with the prior Dutch study [27] on the screening accuracy of the BITSEA Problem scale with regard a CBCL Total Problem score in the clinical range, we found similar results; also a AUC>0.90 and no differences between subgroups. Multiple values for sensitivity and specificity of the BITSEA are reported in the study conducted in the US, because different indicators were used to classify a 'clinical group', and also in the Turkish study, because in their study a range of BITSEA cutpoints was applied. The US-study [19] found comparable mean sensitivity

and specificity for the Problem scale as in our study. However, for the Competence scale in the US-study, a lower sensitivity and slightly higher specificity were found, compared to our study. The Turkish study [26] found slightly higher mean sensitivity and lower mean specificity for the Competence scale, compared to our study. However, the different methods to determine sensitivity and specificity (i.e. different indicators of a 'clinical group' and different methods to determine cutpoints), make it difficult to compare results across these studies.

The Youden index yielded the same cutpoints for boys and girls on the Problem and Competence scales. These results differ from what was found in the US-study [19], where the cutpoints on the Problem scale in children aged 24-29 months differed between boys (score 14) and girls (score 13) and also differed on the Competence scale (girls, score 15; boys, score 14). The Turkish study [26] found the same cutpoint (score 12) on the Competence scale in children aged 24-35 months, for both boys and girls. These differences between studies might be attributed to different characteristics of the study populations. Also, in the Turkish study, the ASD sample size (n=35) was much smaller compared to our ASD sample size (n=159).

The screening accuracy of the newly calculated Autism score was equal for boys and girls, however, the scores with the highest Youden's index differed between boys (score 9) and girls (score 8). Even though the Autism score consists of less items (17 items), its screening accuracy for ASD was better for the total population than the Problem scale (31 items), but not better than the Competence scale (11 items). The Autism score is formally not a separate BITSEA scale and the findings of the present study imply that calculation of the Autism score is unnecessary when the Competence score is known. It was to be expected that the screening accuracy of the Autism score would be at least equally well as the screening accuracy of the Competence scale, since the Autism score consists of 8 of the 11 Competence items. However, the addition of the items from the Problem scale does not further improve the screening accuracy of the Autism score.

Limitations and strengths

Our study has some limitations. First, the BITSEA scores for the ASD-sample are based on BITSEA items that were extracted from the larger pool of ITSEA items, since parents of children in the ASD-sample completed the ITSEA.

Second, as it is expected that children with *typical* development acquire more competencies with age, previous studies have found higher Competence scores in older children, compared to younger children [19,22]. Our community sample consisted of a homogeneous sample with regard to age (M=23.7, SD=0.7). Therefore, it may not be appropriate to generalise our findings on screening accuracy of the Competence scale to children of other ages.

Third, the ASD-sample differed significantly from the community sample with regard to child's gender (more boys), and age (older children). It is likely that these characteristics might have influenced mean BITSEA scale scores; previous studies have found that mean BITSEA scores for boys are less favourable [19,22,23] and that mean Competence scores increase with age [19,22]. Therefore, differences in mean BITSEA scores between the community and ASD-sample might not solely be attributed to the ASD, but also to the demographic characteristics of the samples. To compensate for these differences between conditions, we applied propensity score matching post-hoc. This yielded a sample of 900 matched cases: 750 children in the community sample en 150 in the ASD-sample, with a statistically equal boy/girl ratio (community sample: 74.5% boys, ASD-sample: 80.0% boys). There was still a significant ($p<0.001$) difference between matched cases regarding age (community sample: M=28.9; SD=7.5, ASD-sample: M=31.8; SD=6.4), however the effect size, Cohen's d, was small; 0.38 [39]. We calculated the AUC for the ROC-curves again for the matched sample, and no significant differences (i.e. no overlapping confidence intervals) were found compared to our prior results (data not shown).

Fourth, we do not have follow-up data on the community sample with regard to an ASD diagnosis. However, since the estimated prevalence of ASD is 1% [3,4], we may assume that 31 children out of 3127 children will receive a diagnosis of ASD. It is difficult to estimate exactly what the effect is on our results. However, if the effect would be significant (i.e. a community sample with definitely no children with ASD would lead to other results), the mean BITSEA scores of that community sample would be more favourable than in the present study. This would mean an even larger difference in BITSEA scores, compared to the ASD sample, possibly leading to larger AUC and better sensitivity and specificity than we have found in the present study. So, due to this limitation we rather underestimate than overestimate the 'true' results.

A strength of our study is that the analyses were performed on a large community sample and ASD-sample which adds to the power of the study. Moreover, children in the ASD-sample were diagnosed by experienced clinicians and diagnoses were based on extensive multidisciplinary diagnostic procedures.

Additionally, another strength of our study is that parents completed the questionnaire prior to receiving a diagnostic evaluation. So parents were not biased by knowledge of an ASD diagnosis when answering the questions.

Future research

This study evaluated the screening accuracy of the BITSEA for ASD specifically. We recommend future studies to evaluate the screening accuracy of the BITSEA for a broader range of psychosocial problems.

CONCLUSION

Both the Problem and Competence scales and the Autism score have a good screening accuracy with regard to ASD for the total population and for boys and girls separately. The Autism score does not have added value to the already existing Competence score; for the screening of ASD, the Competence score is just as effective as the Autism score. Furthermore, the BITSEA is a short questionnaire and has in earlier research shown to have good reliability and validity. As mentioned before, in the introduction, early detection instruments that are used in preventive health care should cover a broad range of psychosocial problems. The BITSEA might therefore precede more extensive evaluations on ASD with other instruments, (e.g. the M-CHAT), by more specialized mental health care providers, when scores on the BITSEA indicate concern for ASD. The results of this study indicate that the BITSEA is suitable for use in the setting of (preventive) child health care for the early identification of ASD.

REFERENCES

1. Laurent de Angulo MS, Brouwers-de Jong EA, Bulk A (2005) Ontwikkelingsonderzoek in de jeugdgezondheidszorg. Assen: Koninklijke Van Gorcum.
2. American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.
3. Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, et al. (2009) Prevalence of autism-spectrum conditions: UK school-based population study. British Journal of Psychiatry 194: 500-509.
4. Kogan MD, Blumberg SJ, Schieve LA, Boyle CA, Perrin JM, et al. (2009) Prevalence of Parent-Reported Diagnosis of Autism Spectrum Disorder Among Children in the US, 2007. Pediatrics 124: 1395-1403.
5. Chakrabarti S, Fombonne E (2005) Pervasive developmental disorders in preschool children: Confirmation of high prevalence. American Journal of Psychiatry 162: 1133-1141.
6. Dawson G (2008) Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. Development and Psychopathology 20: 775-803.
7. Seida JK, Ospina MB, Karkhaneh M, Hartling L, Smith V, et al. (2009) Systematic reviews of psychosocial interventions for autism: an umbrella review. Developmental Medicine and Child Neurology 51: 95-104.
8. Dietz C, Swinkels S, van Daalen E, van Engeland H, Buitelaar JK (2006) Screening for autistic spectrum disorder in children aged 14-15 months. II: Population screening with the early screening of autistic traits questionnaire (ESAT). Design and general findings. Journal of Autism and Developmental Disorders 36: 713-722.
9. Filipek PA, Accardo PJ, Ashwal S, Baranek GT, Cook EH, et al. (2000) Practice parameter: Screening and diagnosis of autism - Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. Neurology 55: 468-479.
10. Sandler AD, Brazdziunas D, Cooley WC, de Pijem LG, Hirsch D, et al. (2001) The pediatrician's role in the diagnosis and management of autistic spectrum disorder in children. Pediatrics 107: 1221-1226.
11. Baron-Cohen S, Allen J, Gillberg C (1992) Can autism be detected at 18 months - The needle, the haystack, and the CHAT. British Journal of Psychiatry 161: 839-843.
12. Robins DL, Fein D, Barton ML, Green JA (2001) The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. Journal of Autism and Developmental Disorders 31: 131-144.
13. Sunita, Bilszta JL (2012) Early identification of autism: A comparison of the Checklist for Autism in Toddlers and the Modified Checklist for Autism in Toddlers. Journal of Paediatrics and Child Health.

14. Briggs-Gowan MJ, Carter AS, Bosson-Heenan J, Guyer AE, Horwitz SM (2006) Are infant-toddler social-emotional and behavioral problems transient? Journal of the American Academy of Child and Adolescent Psychiatry 45: 849-858.

15. Egger HL, Angold A (2006) Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. Journal of Child Psychology and Psychiatry 47: 313-337.

16. Scheeringa MS (2001) The differential diagnosis of impaired reciprocal social interaction in children: A review of disorders. Child Psychiatry and Human Development 32: 71-89.

17. Bornstein MH, Hahn CS, Haynes OM (2010) Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: Developmental cascades. Development and Psychopathology 22: 717-735.

18. Yirmiya N, Charman T (2010) The prodrome of autism: early behavioral and biological signs, regression, peri- and post-natal development and genetics. Journal of Child Psychology and Psychiatry 51: 432-458.

19. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.

20. Briggs-Gowan MJ, Carter AS (2002) Brief Infant-Toddler Social and Emotional Assessment (BITSEA) mannual, version 2.0. New Haven: CT: Yale University.

21. Haapsamo H, Ebeling H, Soini H, Joskitt L, Larinen K, et al. (2009) Screening infants with social and emotional problems: A pilot study of the Brief Infant Toddler Social and Emotional Assessment (BITSEA) in Northern Finland. International Journal of Circumpolar Health 68: 386-393.

22. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, et al. (2009) The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behavior & Development 32: 291-297.

23. Kruizinga I, Jansen W, de Haan CL, van der Ende J, Carter AS, et al. (2012) Reliability and validity of the dutch version of the brief infant-toddler social and emotional assessment (BITSEA). PLoS ONE 7.

24. Nunnally JC, Bernstein IH (1994) Psychometric theory. New York: McGraw-Hill.

25. Terwee CB, Bot SDM, de Boer MR, van der Windt D, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 60: 34-42.

26. Karabekiroglu K, Briggs-Gowan MJ, Carter AS, Rodopman-Arman A, Akbas S (2010) The clinical validity and reliability of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). Infant Behavior & Development 33: 503-509.

27. Kruizinga I, Jansen W, Mieloo CL, Carter AS, Raat H (2013) Screening accuracy and clinical application of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). PLoS ONE 8(8): e72602. doi:10.1371/journal.pone.0072602.

28. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington:VT: University of Vermont, Research Center for Children, Youth, and Families.

29. Briggs-Gowan MJ, Carter AS (1998) Preliminary acceptability and psychometrics of the Infant-Toddler Social and Emotional Assessment (ITSEA): A new adult-report questionnaire. Infant Mental Health Journal 19: 422-445.

30. Carter AS, Briggs-Gowan MJ, Jones SM, Little TD (2003) The Infant-Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. Journal of Abnormal Child Psychology 31: 495-514.

31. Briggs-Gowan MJ, Carter AS (2008) Social-emotional screening status in early childhood predicts elementary school outcomes. Pediatrics 121: 957-962.

32. Wet medisch-wetenschappelijk onderzoek met mensen [Dutch Act on research involving human subjects] (1998). Bwb-id:BWBR0009408. http://wetten.overheid.nl/BWBR0009408/geldigheidsdatum_05-02-2010.

33. Oosterling IJ, Wensing M, Swinkels SH, van der Gaag RJ, Visser JC, et al. (2010) Advancing early detection of autism spectrum disorder by applying an integrated two-stage screening approach. Journal of Child Psychology and Psychiatry 51: 250-258.

34. Swinkels SHN, Dietz C, van Daalen E, Kerkhof I, van Engeland H, et al. (2006) Screening for autistic spectrum in children aged 14 to 15 months. I: The development of the early screening of autistic traits questionnaire (ESAT). Journal of Autism and Developmental Disorders 36: 723-732.

35. Oosterling IJ, Swinkels SH, van der Gaag RJ, Visser JC, Dietz C, et al. (2009) Comparative Analysis of Three Screening Instruments for Autism Spectrum Disorder in Toddlers at High Risk. Journal of Autism and Developmental Disorders 39: 897-909.

36. Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.

37. Akobeng AK (2007) Understanding diagnostic tests 3: receiver operating characteristic curves. Acta Paediatrica 96: 644-647.

38. Fischer JE, Bachmann LM, Jaeschke R (2003) A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. Intensive Care Medicine 29: 1043-1051.

39. Cohen J (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

PART TWO

The evaluation of the Brief Infant-Toddler Social and Emotional Assessment

Chapter 6

Evaluation of an early detection tool for social-emotional and behavioral problems in toddlers: The Brief Infant Toddler Social and Emotional Assessment – A cluster randomized trial

Ingrid Kruizinga
Wilma Jansen
Alice S. Carter
Hein Raat

Published as:
Kruizinga I, Jansen W, Carter CS, Raat H (2011) Evaluation of an early detection tool for social-emotional and behavioral problems in toddlers: The Brief Infant Toddler Social and Emotional Assessment - A cluster randomized trial. BMC Public Health 11:494.

ABSTRACT

Background: The prevalence of social-emotional and behavioural problems is estimated to be 8 to 9% among preschool children. Effective early detection tools are needed to promote the provision of adequate care at an early stage. The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) was developed for this purpose. This study evaluates the effectiveness of the BITSEA to enhance social-emotional and behavioural health of preschool children.

Methods and design: A cluster randomized controlled trial is set up in youth health care centers in the larger Rotterdam area in the Netherlands, to evaluate the BITSEA. The 31 youth health care centers are randomly allocated to either the control group or the intervention group. The intervention group uses the scores on the BITSEA and cut-off points to evaluate a child's social-emotional and behavioural health and to decide whether or not the child should be referred. The control group provides care as usual, which involves administering a questionnaire that structures the conversation between child health professionals and parents. At a one year follow-up measurement the social-emotional and behavioural health of all children included in the study population will be evaluated.

Discussion: It is hypothesized that better results will be found, in terms of social-emotional and behavioural health in the intervention group, compared to the control group, due to more adequate early detection, referral and more appropriate and timely care.

Trial registration: Current Controlled Trials NTR2035.

INTRODUCTION

Psychosocial problems, such as social-emotional and behavioural problems, are prevalent among preschool children; in approximately 8-9 percent of preschool children, child health professionals identify psychosocial problems, such as anxious or depressed feelings and aggressive or disobedient behaviour [1,2]. Psychosocial problems in this age group can interfere with everyday functioning [3-5]. Research has demonstrated that problems at preschool age are associated with depressive symptoms, oppositional defiant or conduct disorder, poor peer relationships and social skills, parent- and teacher reported problems with externalizing and internalizing behaviour, poor academic performance and psychiatric problems later in life [6-9]. A retrospective study [10] demonstrated that adolescents with psychosocial problems already had neurocognitive, temperament and behavioural problems at age two or three years old.

It has been recommended that psychosocial problems can be detected at a very young age and followed by appropriate management [11-13]. Research has shown that detection and treatment of psychosocial problems at a young age significantly reduces problems and increases competencies [14,15]. Preventive youth health care, as part of community care or paediatrics, offers an opportunity for the early detection of psychosocial problems among preschool children. Child health professionals, such as physicians and nurses who provide preventive care, may apply early detection of psychosocial problems and if necessary adequate referral or short counselling [16]. In the Netherlands, participation of parents with their child in the preventive youth health care is free of charge and on voluntary basis; almost 95% of the parents of preschool children make use of the youth health care service [17].

Despite the potential impact of psychosocial problems at preschool age and the presence of easily accessible youth health care, only a relatively small number of children with psychosocial problems receive appropriate care [2,5]. One study showed that only 29% of the children with severe problems, based on a Child Behavior Checklist (CBCL) total problem score in the clinical range, were identified by child health professionals [2]. And, in another study, only 13% of the children who scored in the clinical range of the CBCL total problem score were referred to mental health services [5].

In current preventive youth health care in the Netherlands, child health professionals apply a structured questionnaire about psychosocial problems that parents complete before coming to the youth health care center, which helps to structure the conversation between child health professionals and parents; there are no validated norm tables or cut-off scores that indicate when the questionnaire signals a problem [18].

As an alternative to this procedure, it has been recommend to evaluate the use of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) [19] for 1-3 year olds, to detect children at-risk for psychosocial problems and to act upon detection in a coherent, effective way [20].

Objectives

The objective of this study is to evaluate the effectiveness of the BITSEA as an early detection tool for preventive youth health care on children's psychosocial health at one year follow-up, compared to 'care as usual'. Additionally, the feasibility will be evaluated. In this paper we describe the design of this study.

METHODS AND DESIGN

Study design

The design of the study is a cluster randomized controlled trial in which parents of children aged about 24 months are invited to participate. Information on the study is provided to the parents and the parents are asked to provide informed consent. The parents/children are invited by preventive youth health care staff for a regular health check. We identified 31 distinct youth health care centers that were numbered to the purpose of randomization. We use a block randomization so that within each of the 4 organisations, youth health care centers were randomly allocated to either the control group or the intervention group, using random numbers. The child health professionals in the intervention group use the scores on the BITSEA and the cut-off points to assess whether children are at risk for psychosocial problems. The child health professionals in the control group offer usual care by children using a questionnaire for structuring the conversation with the parents. The effect of the intervention will be evaluated after one year of follow-up by comparing CBCL1.5-5 [21] scores between the children in the intervention group and children in the control group, taking into account the baseline measurement on the CBCL1.5-5. The course of the study is presented in Figure 6.1. The Medical Ethics Committee of the Erasmus Medical Centre Rotterdam approved the study protocol (reference number MEC02009-092).

Study procedure and participants

Youth health care centers and randomisation procedure

Four youth health care organizations that consist of a total of 31 distinct youth health care centers that cover both urban and rural areas in the larger Rotterdam area, are participating in this study. Fifteen youth health care centers were randomly allocated to the intervention group; 16 youth health care centers were randomly allocated to the care-as-usual group, applying a block randomisation procedure as described above. Prior to the start of the study, the researchers arranged meetings to explain the study procedure and to instruct the child health professionals in the intervention group about the use and scoring of the BITSEA, with support of a specialized psychiatrist.

Children and their parents

Parents of 3000 children are invited to participate in the study. The study population consists of parents or caregivers of toddlers aged 24 months old at baseline, and 36 months old at one year follow-up. Parents of children in the age range of this study have a high attendance

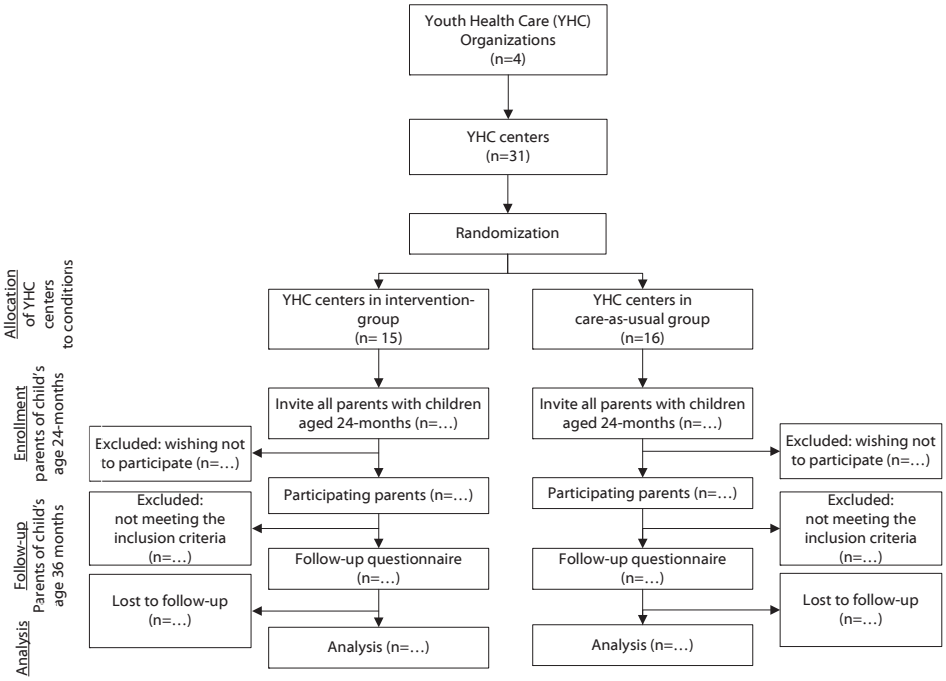


Figure 6.1 Flow chart of the participants and allocation through the trial.

(95%) at the regular health checks at youth health care centers [17]. Children who receive treatment of a mental health professional at baseline will be excluded from the study.

Intervention condition

The 42-item BITSEA is an early detection tool for emotional or behavioural problems and delays in social-emotional competence, including autism spectrum disorders, in toddlers. The BITSEA was developed and applied in the USA, and since then also studied in Turkey and Finland [19,22,23]. It is appropriate for use among children of 12-36 months old and consists of 42 items with 3 response options ('not true/rarely', somewhat true/sometimes', 'very true/often'), that are part of one of 2 multi-item scales, a Problem scale (31 items) or Competence scale (11 items). Per scale the items are summed up into a scale score. In addition to the 42 items, the BITSEA has 2 single-item questions with regard to parents' concerns. Internal consistency of the Problem scale has been reported to be 0.79 and of the Competence scale 0.65 [19]. Ten to 45 day test-retest reliability (ICC) of the Problem scale has been reported to be 0.87 and of the Competence scale 0.85 [19]. Relative to typical parent/teacher agreement [24], the parent/child-care provider correlation was relatively high for the Competence scale (ICC=0.59) and 0.28 for the Problem scale. [19] The BITSEA was translated into Dutch according to international guidelines [25]. Psychometric properties of the BITSEA for the Dutch population will be assessed in the present study.

At the intervention centers child health professionals use the BITSEA [26] as an early detection tool during the regular health check. The child health professionals are trained to score the answers given by parents on the BITSEA and use the cut-off points provided in the literature [19] in their assessment whether children are at risk for or currently experiencing psychosocial problems. Additional information given by parents about timing, duration, intensity of problematic emotions and behaviour is also considered in the assessment of the risk for problems. If, for instance, the problematic emotions/behaviours are mild or are considered to be temporarily, e.g. after a major life event, the child health professional may offer advice about how to cope with the circumstances and may choose to ask the parent back in a few weeks for a follow-up.

The child health professional may choose to refer a child and his/her parents to specialized care when the child is at risk for or currently experiencing problems based on the BITSEA scores and cut-off points, when the problematic behaviours/emotions are severe and not considered to be temporarily. A referral to specialized care is always made after consultation with the physician at the youth health care center.

Control condition

In the control condition youth health care centers at child's age 2 years provide care as usual; i.e. the child health professionals use a short questionnaire [18] that serves as a guide through the conversation between child health professionals and parents. Based on this information the child health professionals may choose to invite parents back for a follow-up visit or to refer to specialized care after consultation with the physician at the youth health care center.

Measurements

Primary outcome measures

The primary outcome of the study is the children's psychosocial health, measured with the Child Behavioral Checklist (CBCL1.5-5) [21]. Child health professionals are blind to this measurement. The 99-item CBCL1.5-5 is designed for children 18 months through 5 years and has two domains (Internalizing and Externalizing Problems and a Total Problem score). Answers are given on a 3-point scale with the following response options: 'not true', 'somewhat or sometimes true' and 'very true or often true'. We apply subclinical and clinical cutpoints for the Dutch population [27]. The primary outcome will be measured at baseline at child's age 24 months and one year after the intervention, at the child's age of 36 months. We hypothesize that children in the intervention group will have a lower Total Problem score on the CBCL1.5-5 at follow-up compared to children in the control group, due to more adequate screening, referral and more appropriate and timely care. For an overview, see Table 6.1.

Secondary outcome measures

A secondary outcome is health related quality of life, i.e. General Health Perceptions subscale and the Growth and Development subscale of the Infant and Toddler Quality of Life Questionnaire (ITQOL) [28,29], measured at follow-up at child's age 36 months. For an overview, see Table 6.1.

Co-variates

Information on parental characteristics (date of birth, ethnicity, immigration characteristics, cultural identity, socio-economic status), children's characteristics (date of birth, sex, ethnicity, day-care attendance, presence of (mental) health problems and treatment for those problems),

Table 6.1 Primary and secondary outcome measures and co-variates in the study

Primary outcome measure	Secondary outcome measure	Co-variates
• CBCL1.5-5 [b,f] *(Total Problem score)*	• ITQOL [f] *(General Health Perceptions subscale)* *(Growth and Development subscale)*	• Date of birth [b] (parents & child) • Sex [b] (child) • Ethnicity [b] (parents & child) • Immigration characteristics [b] (parents) • Cultural identity [b] (parents) • Social economical status [b] (parents) • Day-care attendance [b] (child) • Household composition [b,f] • Major life events [b,f] (parents & child) • Presence of (mental) health [b,f] problems and treatment for those problems (child) • Perceived health of the child rated by parents [f]

[b] measured at baseline (child's age 24 months).
[f] measured at 1 year follow-up (child's age 36 months).

and household composition, major life events and the perceived health of the child rated by parents are obtained from the questionnaires at baseline and at follow-up. For an overview, see Table 6.1.

Statistical analyses

Given the cluster design of the study, multilevel analyses will be applied [30,31]. Linear multilevel analysis will be applied for continuous outcome variables and logistic multilevel analysis for dichotomous outcome variables. Interaction effects of gender and ethnic background with the outcomes will be explored.

Power of the study

Power calculations indicated that a total of 3000 children (and their parents) are needed to detect a difference of 8 points on the CBCL1.5-5 between the control and experimental group, assuming a standard deviation of the CBCL1.5-5 of 26.5 points [32] and an intra-cluster coefficient of 0.1, with a power of 0.80 and alpha 0.05. Assuming a participation of 50% and a lost to follow-up of 30%, we will have complete data at follow-up of 2100 children (1050 in both the intervention and the control group).

Process evaluation

In addition to the effect evaluation, a process evaluation will be carried out, in which both the perspectives of parents and professionals will be taken into account. All parents that are included in this study are asked to evaluate the use of the early detection tool (i.e. level of difficulty, level of understanding, consumed time and satisfaction with the early detection tool as a preparation for the regular health check at the youth health care center). All child health professionals are invited to complete a computer-based process evaluation questionnaire at 6 months after the start of the study. The process evaluation questionnaire consists of items about consumed time, adherence to work instructions, satisfaction with the early detection tool, general perception of the use of the early detection tools in the youth health care and perceived contribution of the early detection tool (a) to the quality of the conversation with parents, (b) to the assessment of the development of the child, and (c) to deciding whether or not to refer.

Furthermore, referrals and consumed care in the year after baseline measurement are assessed at 1-year follow-up, at child's age 36 months; i.e. if a referral to specialized care is made and to which professional; the extent to which parents pursue received referrals, and the diagnosis if one is made are measured at the 1 year follow-up, when children are age 36 months.

DISCUSSION

This paper describes the design of a cluster randomized controlled trial. The trial evaluates the effectiveness of the BITSEA as an early detection tool when used by preventive child health professionals on children's psychosocial health at one year follow-up, compared to 'care as usual'. We hypothesize to find better results, in terms of psychosocial health in the intervention group at one year follow-up, compared to the control group, due to more adequate early detection, referral and more appropriate and timely care.

Strengths of the study are the cluster randomized controlled design, the power of the study, and the setting of the study, which is the daily practice of regular health checks at the youth health care centers that are highly attended by parents. The one year follow-up measurement allows evaluation of the medium term effect of the BITSEA. The study sample will include families with a non-Dutch background, which we expect will add to the generalizability of the results.

Because the study relies primarily on self-report by parents, misclassification might occur. Parents might provide socially desirable answers, e.g. by understating problems or overstating competencies.

A limitation of the study is that the questionnaires are only available in Dutch. For this reason it might be possible that parents with a relatively low level of knowledge of the Dutch language will have some difficulty with the completion of the early detection tool. However, parents have the opportunity to ask for help regarding this issue at the youth health care centers. Furthermore we assess the extent in which parents have understood the questions in the early detection tool as a process measure.

In conclusion, the study evaluates the effectiveness of the BITSEA as an early detection tool to be applied by child health professionals, with the purpose of promoting children's psychosocial health at one year follow-up, compared to 'care as usual'.

ACKNOWLEDGEMENTS

We express our gratitude to J.C. Visser, S. Smeekens and J.M.A. Riksen-Walraven for the forward translation and I. Flink and Z. Op de Macks for the back translation of the BITSEA.

REFERENCES

1. Velderman M, Crone M, Wiefferink C, Reijneveld S (2010) Identification and management of psychosocial problems among toddlers by preventive child health care professionals. European Journal of Public Health 20: 332-338.
2. Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP (2004) Identification and management of psychosocial problems among toddlers in Dutch preventive child health care. Archives of Pediatrics & Adolescent Medicine 158: 811-817.
3. Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP (2001) Identification and management of psychosocial problems by preventive child health care. Archives of Pediatrics & Adolescent Medicine 155: 462-469.
4. Briggs-Gowan MJ, Carter AS, Skuban EM, Horwitz SM (2001) Prevalence of social-emotional and behavioral problems in a community sample of 1-and 2-year-old children. Journal of the American Academy of Child and Adolescent Psychiatry 40: 811-819.
5. Verhulst FC, van der Ende J (1997) Factors associated with child mental health service use in the community. Journal of the American Academy of Child and Adolescent Psychiatry 36: 901-909.
6. Briggs-Gowan MJ, Carter AS (2008) Social-emotional screening status in early childhood predicts elementary school outcomes. Pediatrics 121: 957-962.
7. Campbell SB, Spieker S, Burchinal M, Poe MD (2006) Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. Journal of Child Psychology and Psychiatry 47: 791-800.
8. Hakman M, Sullivan M (2009) The Effect of Task and Maternal Verbosity on Compliance in Toddlers. Infant and Child Development 18: 195-205.
9. Mesman J, Koot HM (2001) Early predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. J Am Acad Child Adolesc Psychiatry 40: 1029-1036.
10. Moffitt TE, Caspi A (2001) Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. Development and Psychopathology 13: 355-375.
11. American Academy of Pediatrics. Committee on children with disabilities: development surveillance and screening of infants and young children. (2001) Pediatrics 108: 192-196.
12. Bricker D, Schoen Davis M, Squires J (2004) Mental health screening in young children. Infants and Young Children 17: 129-144.
13. (2003) Horen, zien, niet zwijgen [See all, hear all, say something]. Inspectie voor de Gezondheidszorg, Inspectie van het Onderwijs, Inspectie Openbare Orde en Veiligheid, Inspectie Jeugdhulpverlening, Jeugdbescherming.
14. Durlak JA, Wells AM (1998) Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. American Journal of Community Psychology 26: 775-802.

15. Elliot J, Prior M, Merrigan C, Ballinger K (2002) Evaluation of a community intervention programme for preschool behaviour problems. Journal of Paediatrics and Child Health 38: 41-50.

16. Basistakenpakket Jeugdgezondheidszorg 0-19 jaar [Youth Health Care Tasks 0-19 years]. (2002). Ministerie van Volksgezondheid, Welzijn en Sport.

17. (2009) Centraal Bureau voor de Statistiek [Statistics Netherlands]. Statistics Netherlands.

18. Kousemaker NPJ (1996) Zoeken, vinden, zorgen delen: de ontwikkeling van een praktijkparadigma voor onderkenning en pedagogische preventie van psychosociale problematiek in de Jeugdgezondheidszorg [Searching, finding, sharing care; the development of a practise paradigm for the descernment and pedagogical prevention of psychosocial problems in Youth Healthcare].

19. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.

20. Hermans J, Ory F, Schrijvers G (Julius Centrum: Utrecht; 2005) Supporting development and parenting: sooner, faster and better: an advice for early detection and interventions in regard to developmental and parenting problems.

21. Achenbach TM, Edelbrock C, Howell CT (1987) Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. J Abnorm Child Psychol 15: 629-650.

22. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, et al. (2009) The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behavior & Development 32: 291-297.

23. Haapsamo H, Ebeling H, Soini H, Joskitt L, Larinen K, et al. (2009) Screening infants with social and emotional problems: A pilot study of the Brief Infant Toddler Social and Emotional Assessment (BITSEA) in Northern Finland. International Journal of Circumpolar Health 68: 386-393.

24. Achenbach TM, McConaughy SH, Howell CT (1987) Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity Psychological Bulletin 101: 213-232.

25. Wild D, Grove A, M Martin, al. e (2005) Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 8: 94-104.

26. Briggs-Gowan MJ, Carter AS (2002) Brief Infant-Toddler Social and Emotional Assessment (BITSEA) mannual, version 2.0. New Haven: CT:Yale University.

27. Verhulst FC, Van der Ende J, Koot HM (1997) Handleiding voor de Teacher's Report Form (TRF). [Manual for the Theacher's Report Form (TRF)]. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

28. Raat H, Landgraf JM, Oostenbrink R, Moll HA, Essink-Bot ML (2007) Reliability and validity of the Infant and Toddler Quality of Life Questionnaire (ITQOL) in a general population and respiratory disease sample. Quality of Life Research 16: 445-460.

29. Raat H, Monhangoo AD, Grootenhuis MA (2006) Pediatric health-related quality of life questionnairese in clinical trials. Current Opinion in Allergy and Clinical Immunology 6: 180-185.

30. Campbell MK, Elbourne DR, Altman DG, Grp C (2004) CONSORT statement: extension to cluster randomised trials. British Medical Journal 328: 702-708.

31. Twisk JWR (2006) Applied Multilevel Analysis: A practical guide. Cambridge: Cambridge University Press.

32. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington: VT: University of Vermont, Research Center for Children, Youth, and Families.

Chapter 7

The effectiveness of the BITSEA as a tool to early detect psychosocial problems in toddlers, a cluster randomized trial

Ingrid Kruizinga
Wilma Jansen
Nicolien C. van Sprang
Alice S. Carter
Hein Raat

Submitted for publication.

ABSTRACT

Objective: Effective early detection tools are needed in child health care to detect psychosocial problems among young children. This study aimed to evaluate the effectiveness of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA), in reducing psychosocial problems at one year follow-up, compared to care as usual.

Method: Well-child centers in Rotterdam, the Netherlands, were allocated in a cluster randomized controlled trial to the intervention condition (BITSEA - 15 centers), or to the control condition ('care-as-usual' - 16 centers). Parents of 2610 2-year-old children (1207 intervention; 1403 control) provided informed consent and completed the baseline and 1-year follow-up questionnaire. Multilevel regression analyses were used to evaluate the effect of condition on psychosocial problems and health related quality of life (i.e. respectively Child Behavior Checklist and Infant-Toddler Quality of Life). The number of (pursuits of) referrals and acceptability of the BITSEA were also evaluated.

Results: Children in the intervention condition scored more favourably on the CBCL at follow-up than children in the control condition: B=-2.43 (95% confidence interval [95%CI] = -3.53;-1.33 $p<0.001$). There were no differences between conditions regarding ITQOL. Child health professionals reported referring fewer children in the intervention condition (n=56, 5.7%), compared to the control condition (n=95, 7.9%; $p<0.05$). There was no intervention effect on parents' reported number of referrals pursued. It took less time to complete (parents) or work with (child health professional) the BITSEA, compared to care as usual. In the control condition, 84.2% of the parents felt (very) well prepared for the well-child visit, compared to 77.9% in the intervention condition ($p<0.001$).

Conclusion: The results support the use of the BITSEA as a tool for child health professionals in the early detection of psychosocial problems in 2-year-olds. We recommend future studies in large and varied populations to replicate these findings.

Trial registration: Current Controlled Trials NTR2035.

INTRODUCTION

The prevalence of psychosocial problems, such as behavioural and emotional problems, is relatively high in preschool children [1-3]; child health professionals identify psychosocial problems in 7-25% of the preschool children [2-5]. It is important to detect problems at an early stage, since the identification and treatment of psychosocial problems at a young age may reduce problems and increases competencies at later ages [6,7]. In the Netherlands, the health care system offers publicly funded preventive programs for all children from birth to the age of 19 years. This setting offers an excellent opportunity to detect psychosocial problems early.

A feasible approach for facilitating early detection of psychosocial problems is to use parent-completed questionnaires as part of routine primary care visits (i.e. well-child visits) [8]. Early detection instruments for psychosocial problems, intended for use in preventive child health care, should have adequate psychometric properties, and should also be short, easy to administer, score and interpret [9,10]. Furthermore, early detection instruments that are used in a public health care setting should cover a broad range of psychosocial problems, since limited time and capacity for the well-child visits make it undesirable to screen for each psychosocial problem separately. Also, it has been shown that psychosocial problems tend to co-occur [11,12], and that individual problems may apply to more than one disorder [13].

The Child Behavior Checklist 1.5-5 (CBCL1.5-5) [14] and Infant-Toddler Social and Emotional Assessment (ITSEA) [15,16] are early detection instruments that are well-validated and measure a broad range of psychosocial problems, and in the case of the ITSEA also delays in competencies. However both instruments are too extensive to apply in the context of well-child visits. Short comprehensive instruments that are appropriate to measure psychosocial problems in children of preschool age are limited [17]. Existing instruments, such as the Eyberg Child Behavior Inventory [18] or the Toddler Behavior Screening Inventory [19], only measure problem behaviour and do not address social-emotional competencies. Measuring delays in social-emotional competence, however, is also important since delays in competence are for instance related to internalising and externalising problems later in life [20]. There remains a need for a short instrument that measures both problems and delays in competence.

The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) [21] is a short (42 items) questionnaire, that measures both problems (Problem scale) and delays in the acquisition of competencies (Competence scale) in 1-3 year olds, and also consists of items designed to measure symptoms of autism spectrum disorders. Several studies, conducted among a

large (N=3127) and diverse community sample in the preventive child health care of the Netherlands, have indicated that the psychometric properties of the BITSEA are acceptable to good [22]. An adequate Cronbach's alpha (i.e. >0.70 [23]) was found for the Problem scale (0.76) and marginal for the Competence scale (0.63). Test-retest reliability was adequate (>0.70 [24]) for the Problem scale (0.75) and marginal for the Competence scale (0.61). The BITSEA Problem scale was positively correlated with the CBCL, Pearson coefficients of 0.66 (Internalizing), 0.65 (Externalizing) and 0.75 (Total Problem). The BITSEA Competence score was negatively correlated with the CBCL, Pearson coefficients of -0.26 (Internalizing), -0.23 (Externalizing) and -0.26 (Total Problem). All correlations were significant ($p<0.01$). The mean BITSEA score was compared between a group of parents that worried about the development of their child and a group that did not worry. The Problem and Competence score were significantly less favourable in the group of parents that worried, compared to the group of parents that did not worry (effect sizes were respectively 0.93 and 0.52). Additionally, the BITSEA is able to discriminate between children with and without psychosocial problems; the Problem scale sensitivity is 0.83 and the specificity is 0.84; the Competence scale sensitivity is 0.95 and the specificity is 0.90 (with respectively reference groups of children with a CBCL Total Problem score >60, and children diagnosed with autism spectrum disorder) [25,26]. These results are confirmed by studies in other countries [21,27-31].

Objective of the study

In the present study, we evaluated the effectiveness of the BITSEA, a questionnaire that supports child health professionals in detecting psychosocial problems in 2-year olds, in psychosocial problems and health related quality of life at child age 3-years (at one year follow-up), compared to care as usual. In care as usual (i.e. control condition), the KIPPPI (KIPPPI is a Dutch acronym for Brief Instrument Psychological and Pedagogical Problem Inventory) is used [32].

The research questions in this study were:

1. Are there fewer parent reported psychosocial problems at follow-up (i.e. lower Child Behavior Checklist [CBCL] Total Problem score) in the intervention group, compared to care as usual?
2. Is health related quality of life at follow-up better (i.e. higher Infant-Toddler Quality of Life [ITQOL], Growth and Development, and General Health Perceptions scores) in the intervention group, compared to care as usual?

Additionally, we explored whether there is a difference between conditions in number of referrals and pursuits of referrals, and whether there is a difference between conditions in the acceptability as perceived by parents and child health professionals.

METHODS

Ethics statement

The Medical Ethical Committee of the Erasmus Medical Center has reviewed the study proposal (see File Protocol S1) and granted permission to conduct the study (MEC-2009-092, February 3rd 2009). They decided that the study does not fall within the ambit of the Medical Research Involving Human Subjects Act ('WMO'). All parents who participated in the study provided written informed consent. We are prepared to make the data available upon request.

Study design

The present study is a cluster randomized controlled trial, conducted in child health care centers in the larger Rotterdam area in the period April 2010-April 2012. Details of our study design were published previously [33] (see File S1). Child health care centers were randomly allocated (by researcher IK) to the control group (16 child health centers in total) or to the intervention group (15 child health centers in total), stratified by organization, using random numbers. See Figure 7.1.

At baseline, parents of children aged 24 months old were invited by child health care organizations for a well-child visit. Along with this invitation, parents received information regarding this study and a child health monitor questionnaire, including among others the CBCL and either the BITSEA or KIPPPI. The BITSEA or KIPPPI was used during the well-child visit by the child health professional to assess the psychosocial development of the child. Child health professionals were not blind to the answers on the CBCL, however they did not score the CBCL and they were instructed not to use the answers that parents provided on the CBCL in their assessment of the psychosocial development of the child. Prior to the start of the study, the child health professionals in the intervention group were trained in the use and scoring of the BITSEA. Instructions were provided by the researchers and a specialized psychiatrist.

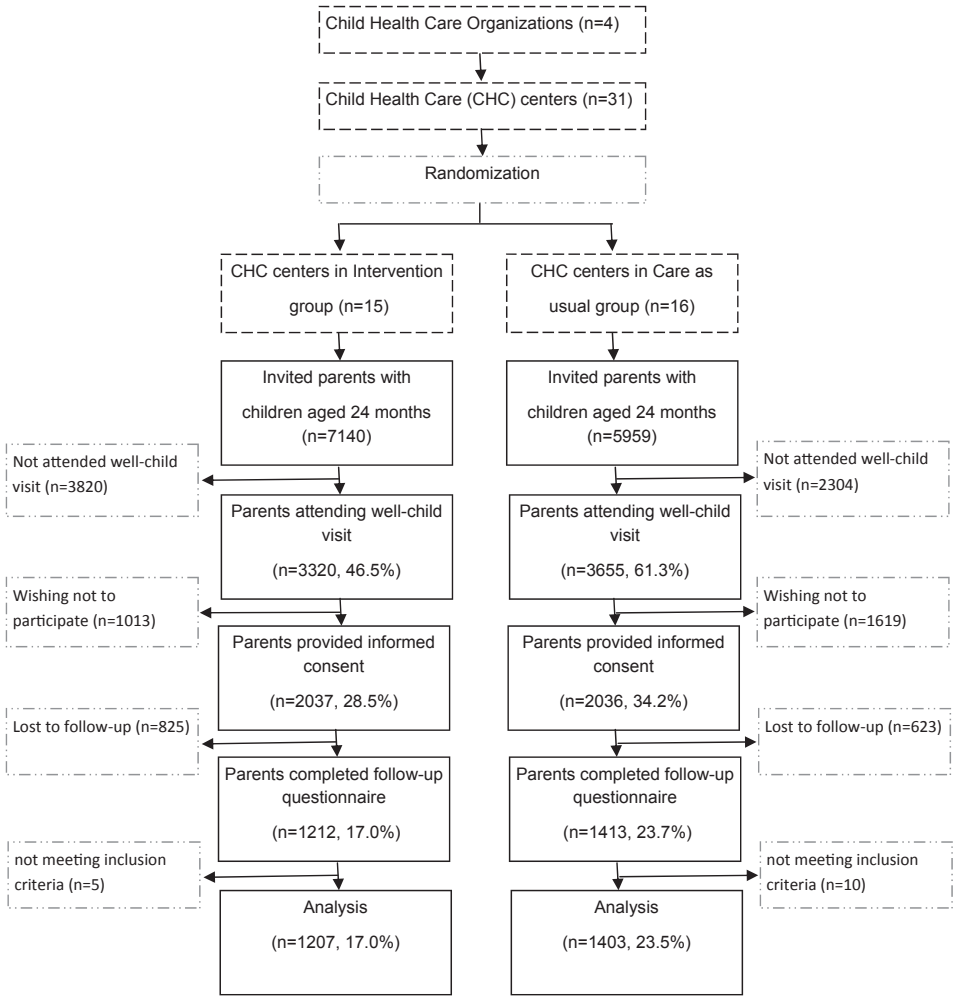


Figure 7.1 Flowchart of the participants through the study.

At one-year follow-up (child aged 36 months), parents who had agreed to participate in the study and provided informed consent, were sent a follow-up questionnaire (with a maximum of 2 reminders), containing the CBCL, ITQOL, either the BITSEA or the KIPPPI and questions regarding pursuit of referrals in the previous year.

Blinding: It was not possible to conduct a double blind study; both parents and child health care providers were aware of the type of questionnaire that was used. Some child health professionals may have been aware that the BITSEA was recommended as 'a promising questionnaire' by an advisory committee to the Dutch government [34], which could have

influenced their assessment of the child and subsequent (referral) decisions. Also, in this study, it was impossible to keep members of the study team blind.

Treatment contamination: It is not plausible that many parents were aware of the BITSEA being possibly better than care as usual, so we do not expect that being assigned to either condition influenced the responses given on the questionnaires. Also, it is not plausible that parents chose to go to another child health care center, outside of the neighborhood where they live, because they preferred the other instrument. For professionals in each of the condition groups it was not feasible to use the other instrument. So, the presence of treatment contamination was not likely.

The protocol for this trial and supporting CONSORT checklist are available as supporting information; see Checklist S1 and Protocol S1.

Study population

Child health care centers

Four child health care organizations consisting of 31 distinct child health centers (i.e. the clusters) that cover both urban and rural areas in the larger Rotterdam area, participated in this study. The characteristics of the clusters per condition are: **Intervention condition:** M_{size}=78.0 (range mean=24-157), mean SD_{size}=55.29; M_{CBCL}=17.8 (range mean=13.5-22.3), mean SD_{CBCL}=3.00; M_{age}=23.7 (range mean=23.25-24.18), mean SD_{age}=0.24; mean percentage boys=49.1 (range mean=34.2-66.7); mean percentage native=75.5 (range mean=47.3-89.2). **Control condition:** M_{size}=80.0, mean SD_{size}=73.92; M_{CBCL}=19.1 mean SD_{CBCL}=2.37; M_{age}=24.0 (range mean=23.2-24.9) mean SD_{age}=0.26; mean percentage boys=51.2 (range mean=30.0-63.9); mean percentage native=79.4 (range mean=62.7-90.0).

Children and their parents

For a flowchart of the participants through the trial, see Figure 7.1. Parents of 13099 2-year old children were invited to participate in the study [7140 (54.5%) intervention; 5959 (45.5%) control]. Parents of 6975 (53.2%) children [3320 (47.6%) intervention; 3655 (52.4%) control] attended the well-child visit. Parents of 4073 (31.1%) [2037 (50.0%) intervention; 2036 (50.0%) control] provided informed consent. Parents of 2625 (20.0%) children [1212 (46.2%) intervention; 1413 (53.8%) control] also completed the follow-up questionnaire. Children were excluded from analyses if they had missing data on gender and ethnicity, and if there were too many missing items on the questionnaires (BITSEA Problem scale>5

items missing; BITSEA Competence scale>2 items missing; KIPPPI>25% items missing per scale). None of the children were under treatment of a mental health professional during the time of inclusion. After exclusion, 2610 (19.9%) children [1207 (46.2%) intervention; 1403 (53.8%) control] remained for analyses.

Drop out was for the intervention and control condition, respectively 59.2% and 68.9% of the parents who attended the well-child visit and provided informed consent. Drop-out was higher among younger parents with a lower education, parents and children with a non-native Dutch ethnicity and older children (measured at baseline, $p<0.05$). However, effect sizes (Cohen's d and Cramer's V) for these differences were low and ranged from 0.01-0.22. There was no difference between children included in the study and drop-out in gender and CBCL Total Problem score at baseline. For the CBCL Total Problem score and for variables that differed between drop-outs and participants in the study and, we evaluated with logistic regression whether there was a difference between conditions in drop-outs. None of the odds ratios were significant ($p>0.05$). See Figure 7.1 for a flow-chart of the participants through the study. No data was available on the group of parents who did not attend the well-child visit. Therefore, evaluating the differences between parents included in the study and parents who did not attend the well-child visit (and thus not participated in the study), was not possible.

Power calculation

As reported in the design paper on this study [33], power calculations indicated that, with 30 clusters, a total of 2100 children (and their parents) would be needed to detect a difference of 8 points on the CBCL1.5-5 between the control and intervention condition, assuming a standard deviation of the CBCL1 Total Problem score of 26.5 points [14] and an intra-cluster coefficient of 0.1, with a power of 0.80 and alpha 0.05.

Intervention condition

In the intervention condition, child health providers used the BITSEA completed by parents as an early detection tool for emotional and behavioural problems. The 42-item BITSEA is an early detection instrument for psychosocial problems and delays in competence, including autism spectrum disorders, in 12 to 36 month old children. Thirty-one items sum up to a Problem score (a high score is less favourable) and 11 items sum up to a Competence score (a low score is less favourable). Studies have shown that the BITSEA has good psychometric

properties [21,22,27,29]. At baseline, child health providers calculated the Problem and Competence score of the BITSEA. To assess the psychosocial development of the child, cutpoints as identified in an earlier study [35] were used. Children with a BITSEA score above the cutpoint(s) warranted a conversation with the parents regarding the 'at risk behaviour', more specifically; level of worry, frequency and intensity of the behaviour, timing and onset of the behaviour, and the context in which the behaviour is manifested and the cultural meaning of the behaviour to the parent were discussed [35]. The child health professionals received additional training prior to the implementation of the BITSEA, so that they were able to objectively score the BITSEA and they had the appropriate skills to be able to have a conversation with the parents if an at risk score on the BITSEA was present, in order to gain more insight in the significance of the problems. The training was provided by the researchers and a child psychiatrist, and consisted of a presentation on psychosocial problems (prevalence, signals that indicate psychosocial problems, importance of early detection, worry of parents, conversation with parents during the well-child visit); background information of the BITSEA; and practical scoring instruction with the opportunity to practice. Also, the professionals received a manual and a scoring aid. The CHP made referrals to other health care providers in consultation with the parent when the 'at risk behaviour' seemed to be clinically significant. Follow up appointments were made if clarification of problems were necessary or for consultation on less serious problems that did not need referral.

Control condition

In the control condition, care as usual was offered during the well-child visit, i.e. child health professionals used the parent completed KIPPPI as an early detection tool. The 67 item KIPPPI measures psychosocial problems, which might be possible pedagogical challenges for the parents. The KIPPPI consist of a Wellbeing scale (31 items, describing problems related to eating/drinking; sleeping; activity; mood; behaviour), Competence scale (25 items, describing problems related to cognitive development; language; play; contact), and an Autonomy scale (11 items, describing problems related to toilet training; motor skills; independence). The KIPPPI Total score is the sum of the scale scores (high scores on the KIPPPI are less favourable), and was calculated in a previous study to be able to evaluate the psychometric properties of the KIPPPI [36]. Studies have shown that the KIPPPI has adequate to good reliability and validity [36,37]. However, in care as usual the KIPPPI is not scored by the child health professional; the KIPPPI serves as a guide in the assessment of the development of the child and its use therefore relies on the subjective interpretation by the child health professional

of the answers given on the KIPPPI. The reason that the KIPPPI was not scored in care as usual is that no empirically determined cutpoints exist. No additional training of the child health professionals was provided. At baseline, the child health professional discusses the items with high ratings (indicating a problem) with parents and assesses whether the difficulties stem from a problem in the child (i.e. psychosocial), or the parents (i.e. pedagogical), or the parent-child interaction. Based on this information the child health professionals may choose to refer parents and their child to other health care providers. Follow up appointments were made if clarification of problems were necessary or for consultation on less serious problems that did not need referral.

Measures

Effect evaluation

The primary outcome measure in the present study is parent reported psychosocial problems of the children participating in the study at one year follow-up. In both the intervention and control condition, parents completed the CBCL1.5-5 at baseline and at follow-up. The well-validated [14] 100-item CBCL is designed for children aged 18-months to 5-years and has two domains (Internalizing and Externalizing) and provides a Total Problem score. A higher CBCL Total Problem score is less favourable. In this study, the Total Problem score (raw score) is used to measure psychosocial problems.

The secondary outcome measure in the present study is parent reported health related quality of life of the children participating in the study at one year follow-up. In both intervention and control condition, parents completed two scales of the ITQOL; Growth and Development, and General Health Perceptions. The ITQOL measures the health related quality of life of children between 2 months and 5 years old [38,39]. Items were reversed scored (if appropriate) so that a higher score is more favourable.

Referrals and acceptability

In addition to the evaluation of the effect on parent reported psychosocial problems and health related quality of life, we also evaluated referrals and acceptability. Referrals were evaluated as a possible explanation for when a difference in outcome variables was found between conditions, and was defined as an actual referral to a mental health professional and/or the advice for a follow-up consult. Referrals were evaluated as registered by the child health professional at baseline on a separate registration form or in a digital medical record system depending on

the automatization level of the child health organization involved, Additionally, the extent to which parents pursued received referrals was evaluated by parental report at follow-up.

Acceptability was evaluated from two perspectives; 1) parents evaluated at baseline the child health monitor questionnaire that included either the BITSEA or KIPPPI on the following aspects: a) perceived duration (dichotomized as '(much) too short/exactly good-(much) too long'); b) actual duration (dichotomized as 'shorter than 15 minutes-longer than 15 minutes'); c) preparation for the well-child visit (dichotomized as 'not good (at all)/mediocre-(very) good'); and 2) 105 (92.9%) child health professionals completed, about a half year after the start of the study, an electronic questionnaire regarding the perceived acceptability of the questionnaires. The following aspects were evaluated: a) time spent on the questionnaire before the well-child visit (i.e. scoring of the BITSEA or looking at the answers on the KIPPPI); b) time spent on the questionnaire during the well-child visit (i.e. discussing the scores/answers with the parents; c) supportive in detecting psychosocial problems (binary; yes or no); d) supportive in the conversation with the parents (binary; yes or no); e) supportive in the assessment of the development of the child (binary; yes or no).

Demographic variables

Demographic variables were assessed in the baseline questionnaire: parental age, country of birth, and educational level and child gender and ethnicity. A child was considered native if both parents were born in the Netherlands, according to Statistics Netherlands [40].

Analyses

Descriptive statistics were used to describe the characteristics of the parents and children in the two conditions. Differences between the intervention and control condition, as measured at baseline, were tested with an independent t-test (continuous variables) and Chi-square test (categorical variables). Effect sizes of the differences between intervention and control condition were calculated. For the continuous variables Cohen's d was calculated: *Cohen's d=[mean1–mean2]/SD1* and is interpreted as follows: $0.20 \leq d < 0.50$ indicates a small effect, $0.50 \leq d < 0.80$ indicates a medium effect and $d \geq 0.80$ indicates a large effect.[41]. For the categorical variables Cramers V [42] was calculated: $V=(chi^2/(\text{sample size}(\text{smallest value of number of columns or rows})))^{1/2}$. For the interpretation of Cramers V, the following characterization is applied <0.10 low association; 0.10-0.25 moderate association; >0.25 high association;

Effect evaluation

The multilevel regression analysis (with child health centers as clusters) is performed with psychosocial problems at follow-up as primary outcome measure (i.e. CBCL Total problem score at follow-up) and with health related quality of life at follow-up as secondary outcome measure (i.e. ITQOL Growth and Development scale, and General Health Perceptions scale), and the intervention condition as independent variable (i.e. BITSEA or KIPPPI). The multilevel regression analyses were adjusted for psychosocial problems at baseline (i.e. CBCL Total problem score at baseline), child gender and ethnicity, because previous studies show that gender and ethnicity are associated with psychosocial problems [43-45]. Additionally, when a significant intervention effect was found, the analyses will be performed with correction for all variables that differed between conditions at baseline. Interaction effects for child gender and condition, and child ethnicity and condition on psychosocial problems and health related quality of life were explored when the main effect of condition was significant.

The intracluster correlation coefficient (ICC) is calculated as $\rho=S^{2b}/(S^{2b}+S^{2w})$, where $S^{2b}$is the variance between the cluster and S^{2w} is the variance within clusters. An ICC of less than 0.1 is considered small [46]. The ICC in this study is; $\rho=3.2597/(254.41+3.2597)=0.01$.

Referrals and acceptability

Differences between the intervention and control condition in number of referrals, pursuits of referrals and acceptability were tested with an independent t-test (continuous variables) and Chi-square test (categorical variables). When these tests indicated a significant difference between conditions, an odds ratio was calculated.

Multilevel regression analysis were performed in SAS 9.3 (SAS Institute Inc., 2011), all other analyses were performed in SPSS 21.0 (SPSS Inc., 2012).

RESULTS

Demographic characteristics of the parents and children are presented in Table 7.1, as well as the mean questionnaire scores at baseline. Participants in the intervention condition differed ($p<0.05$) in the following aspects from the control condition (see Table 7.1), mean age of the mother and father was higher (effect size is for both parents 0.09); fewer mothers and fathers were born in the Netherlands (effect size is respectively 0.12 and 0.14); more mothers and fathers had attended vocational education or university (effect size is respectively 0.10 and

Table 7.1 Baseline demographic characteristics and mean questionnaire scores for the intervention and control condition, N=2610

	Intervention condition BITSEA, n=1207	Control condition KIPPPI, n=1403	Effect size
Mother characteristics			
Mean (SD) age	34.29 (4.69)*	33.88 (4.67)*	0.09[a]
Country of birth [n (%) Dutch]	920 (76.22)**	1198 (85.39)**	0.12[b]
Educational level [n (%) higher vocational/university]	631 (54.30)**	592 (43.98)**	0.10[b]
Father characteristics			
Mean (SD) age	36.80 (5.29)*	36.35 (5.15)*	0.09[a]
Country of birth [n (%) Dutch]	892 (73.90)**	1199 (85.46)**	0.14[b]
Educational level [n (%) higher vocational/university]	571 (51.63)**	524 (40.03)**	0.12[b]
Child characteristics			
Gender [n (%) boys]	594 (49.20)	733 (52.20)	n.a.
Ethnicity [n (%) native Dutch]	849 (70.34)**	1139 (81.18)**	0.13[b]
Mean (SD) age	23.69 (0.69)**	23.98 (1.09)**	0.27[a]
Mean (SD) CBCL Total Problem score	18.87 (15.31)*	20.32 (15.66)*	0.09[a]
Mean (SD) BITSEA Problem score	7.38 (4.85)	n.a.	n.a.
Mean (SD) BITSEA Competence score	17.90 (2.93)	n.a.	n.a.
Mean (SD) KIPPPI score	n.a.	40.93 (14.16)	n.a.

[a] Cohen's d (Mean1-Mean2/SD1), is considered small.
[b] Coefficient Phi, is considered small.
* Significant difference, p<0.05.
** Significant difference, p<0.01.

0.12). Compared to the control condition, the intervention condition consisted of fewer native Dutch children (effect size=0.13). Children in the intervention condition were younger (effect size=0.27) and scored less favourably on the CBCL Total Problem score at baseline (effect size=0.09). Effect sizes of all differences in variables between intervention and control condition were small [41]. We have performed the analyses corrected for all the variables that differed between intervention and control condition, and the results were similar as those reported (data not shown).

Effect evaluation

Psychosocial problems

As presented in Table 7.2, children in the intervention condition had lower mean CBCL Total Problem scores compared to children in the control condition both at baseline (respectively M=18.87, SD=15.31 and M=20.32, SD=15.66) and follow-up (respectively M=19.32,

Table 7.2 Mean (SD) CBCL Total Problem score, and ITQOL Growth and Development, and General Health scores (N=2610)

	Intervention, n=1207		Control, n=1403		Difference between conditions
	Baseline	Follow-up	Baseline	Follow-up	Effect size* (baseline;follow-up)
CBCL Total Problem score	18.87 (15.31)[a]	19.32 (15.71)[a]	20.32 (15.66)[ab]	22.51 (16.28)[ab]	0.09;0.20
ITQOL Growth and Development	n.a.	91.39 (11.62)	n.a.	92.11 (10.09)	n.a.;0.07
ITQOL General Health	n.a.	81.77 (14.66)	n.a.	82.43 (14.02)	n.a.;0.05

[a] Significant difference between intervention and control condition in CBLC Total Problem score at baseline and follow-up ($p<0.05$).
[b] Significant difference within control condition between baseline and follow-up CBCL Total Problem score ($p<0.05$).
* Cohen's d (Mean1-Mean2/SD1), is considered small.

SD=15.71 and M=22.51, SD=16.28). In the control condition, mean CBCL Total Problem score was significantly higher at follow-up than at baseline. There was no difference in the intervention condition between baseline and follow-up CBCL Total Problem score ($p>0.05$).

Children in the intervention condition scored more favourably on the CBCL Total Problem score at follow-up than children in the control condition (controlled for CBCL Total Problem score at baseline, child gender and ethnicity); the unstandardized regression coefficient for type of condition was significant ($p<0.001$), B=-2.43 (95% confidence interval [95%CI] = -3.53;-1.33), see Table 7.3. Effect sizes were small: Cohen's d for the difference between conditions in CBCL Total Problem score was at baseline 0.09 and at follow-up 0.20, see Table 7.2.

Similar results were found when the analyses were performed corrected for all the variables that differed between conditions at baseline, see Table 7.4.

Health related quality of life

There were no significant differences ($p>0.05$) between conditions in ITQOL scores on the Growth and Development scale nor on the General Health Perceptions scale, see Table 7.3.

Table 7.3 Regression coefficients and confidence intervals (95%CI) from the multilevel regression models evaluating the association between condition and CBCL Total Problem score and ITQOL scores at follow-up corrected for confounders (N=2230)

	Model 1 beta (95%CI)	Model 2 beta (95%CI)	Model 3 beta (95%CI)	Model 4 beta (95%CI)	Model 5 beta (95%CI)
Primary outcome measure: CBCL Total Problem score at follow-up					
Condition (intervention)	-3.45 (-4.80;-2.10)***	-2.40 (-3.71;-1.09)***	-2.43 (-3.53;-1.33)***	-2.28 (-3.77;-0.78)**	-5.07 (-7.22;-2.92)***
CBCL Total Problem score baseline		0.72 (0.68;0.76)***	0.71 (0.67;0.75)***	0.71 (0.67;0.74)***	0.71 (0.67;0.74)***
Child gender (boy)			1.47 (0.47;2.47)**	1.60 (0.29;2.91)*	1.47 (0.48;2.47)**
Child ethnicity (native)			-0.96 (-2.19;0.27)	-0.96 (-2.19;0.26)	-2.67 (-4.37;-0.96)**
Condition x gender				-0.31 (-2.33;1.70)	
Condition x ethnicity					3.44 (1.05;5.83)
Secondary outcome measure: ITQOL Growth and Development score at follow-up					
Condition (intervention)	-0.66 (-1.58;0.26)	-0.76 (-1.66;0.14)	-0.76 (-1.68;0.16)		
CBCL Total Problem score baseline		-0.12 (-0.16;-0.08)***	-0.12 (-0.16;-0.08)***		
Child gender (boy)			-0.62 (-1.52;0.28)		
Child etnicity (native)			-0.16 (-1.26;0.94)		
Secondary outcome measure: ITQOL General Health score at follow-up					
Condition (intervention)	-0.70 (-2.05;0.65)	-1.01 (-2.24;0.22)	-0.10 (-1.32;1.12)		
CBCL Total Problem score baseline		-0.23 (-0.27;-0.19)***	-0.23 (-0.27;-0.19)***		
Child gender (boy)			-1.43 (-2.59;-0.27)*		
Child etnicity (native)			0.81 (-0.60;2.22)		

*** p<0.001; ** p<0.01; * p<0.05.

Table 7.4 Regression coefficients and confidence intervals (95%CI) from the multilevel regression models evaluating the association between condition and CBCL Total Problem score at follow-up corrected for confounders (N=2230)

	beta (95%CI)
Primary outcome measure: CBCL Total Problem score at follow-up	
Condition (intervention)	-2.32 (-3.75;-0.89)**
CBCL Total Problem score baseline	0.70 (0.66;0.74)***
Child gender (boy)	1.40 (0.34;2.48)**
Child ethnicity (native)	3.39 (-0.40;7.18)

NB. Corrected for parental country of birth; parental educational level; age of the child; parental age.
*** p<0.001; ** p<0.01.

Interactions

Interaction effects were explored. For the primary outcome variable (i.e. CBCL Total Problem score), we evaluated separately condition by gender and condition by ethnicity interactions; both interaction effects were non-significant (p>0.05) (see Table 7.5). For the secondary outcome variable (i.e. ITQOL score on the Growth and Development scale and General Health scale, we did not evaluated interaction effect since the main effect of condition was not significant (p>0.05).

Referrals and acceptability

Referrals

Given the differences between conditions on the primary outcome, we evaluated the number of referrals as a possible explanation. There was a main effect of the intervention (p=0.042) on the number of referrals registered by the child health professionals: fewer children in the intervention condition (n=56, 5.7%) were referred, compared to the control condition (n=95, 7.9%). Children in the control condition were 1.42 times more likely to be referred (odds ratio=1.42, 95%CI=1.01-2.00), compared to children in the intervention condition. There was no effect of the intervention (p>0.05) on the number of parent reported referrals pursued (n=58, 92.1% versus n=54, 91.5%, for the intervention and control groups respectively).

Acceptability

There were significant differences (p<0.001) between the intervention and control condition in the acceptability as perceived by parents. Completion of the BITSEA took less time compared to completion of the KIPPPI: The duration of completion was less than 15 minutes for 592

parents (49.9%) in the intervention condition and for 555 parents (40.6%) in the control condition. The preparation for the well-child visit was considered (very) good by 908 parents (77.9%) in the intervention condition and by 1146 parents (84.2%) in the control condition.

The only significant difference ($p<0.01$) between intervention and control condition in acceptability, as perceived by the child health professionals, was the time that they spent on the questionnaire during the well-child visit. In the intervention condition the mean time was 8.22 minutes (SD=5.59) and in the control condition 12.08 minutes (SD=6.16). See Table 7.6.

Table 7.5 Number (percentages) of referred children in the total sample and in the subsample with baseline 'at risk' scores on the CBCL and BITSEA

	Intervention		Control		
	n	Referred	n	Referred	p
CBCL Total Problem score					
At risk score (score>60)	17 (1.41)	4 (23.5)	21 (1.50)	9 (42.9)	0.212
Low score	944 (78.2)	51 (5.4)	1149 (81.9)	81 (7.0)	0.123
BITSEA Problem scale					
At risk (score≥15 boys/ ≥14 girls)	89 (9.1)	9 (10.1)	n.a.	n.a.	n.a.
Low score	886 (90.9)	47 (5.3)	n.a.	n.a.	n.a.
BITSEA Competence scale					
At risk (score≤14 boys / ≤15 girls)	140 (14.3)	12 (8.6)	n.a.	n.a.	n.a.
Low score	839 (85.7)	44 (5.2)	n.a.	n.a.	n.a.
Total	983 (81.4)	56 (5.7)	1200 (85.5)	95 (7.9)	0.042*

* Significant differences in number of referrals between intervention and control condition, as reported by child health professionals.

Table 7.6 Differences between intervention and control condition regarding acceptability

	Intervention BITSEA	Control KIPPPI
Acceptability parents - N=2610		
Perceived duration - (too)short/exactly good [n (%)]	820 (69.7)	924 (67.6)
Actual duration - <15 minutes [n (%)]	592 (49.9)***	555 (40.6)***
Preparation for well-child visit - (very) good preparation [n (%)]	908 (77.9)***	1146 (84.2)***
Acceptability child health professional - n=105		
Time spend before consult [M (SD)]	4.85 (3.69)	3.95 (2.44)
Time spend during consult [M (SD)]	8.22 (5.89)**	12.08 (6.16)**
Supportive detecting problems [n (%)]	31 (72.1)	31 (64.6)
Supportive conversation with parents [n (%)]	29 (67.4)	39 (83.0)
Supportive assessment development [n (%)]	25 (62.5)	20 (43.5)

*** Significant difference, $p<0.001$; ** significant difference, $p<0.01$; * significant difference, $p<0.05$.

DISCUSSION

In the present study, we evaluated the effect at one year follow-up of the use by preventive child health care of the BITSEA as a questionnaire for the early detection of parent reported psychosocial problems in 2-year olds, compared to care as usual; the KIPPPI. Furthermore, we assessed the number of referrals, number of pursuits of referrals and acceptability of the BITSEA and KIPPPI as perceived by parents and professionals. The results indicate that children in the intervention condition had fewer psychosocial problems at follow-up (i.e. lower CBCL Total Problem score), compared to the control condition. No intervention effects were observed regarding health related quality of life. Child health professionals registered more referrals in the control condition than in the intervention condition. The BITSEA takes less time for parents to complete and for the child health professional to score, compared to the KIPPPI. Parents perceived the KIPPPI as a better preparation for the well-child visit, compared to the BITSEA.

Effect evaluation and referrals

The difference in CBCL Total Problem score between conditions at follow-up is significant, but the effect size is small. A large effect might not be expected given the relatively long follow-up period of one year in which many variables can influence the psychosocial well-being of a child. Post-hoc we evaluated the subscales of the CBCL, to see what subscale (if any) caused the differences in the Total Problem score. We performed an independent t-test and the significant subscales were: Emotionally Reactive ($p=0.009$); Anxious/Depressed ($p=0.001$); Somatic Complaints ($p<0.001$); Withdrawn ($p<0.001$); Aggressive Behavior ($p=0.001$); Other ($p<0.001$); Internal ($p<0.001$); External ($p<001$). There is not one specific subscale responsible for the difference between conditions in CBCL Total Problem score, but several. One might expect that children who are referred will have fewer psychosocial problems one year later, due to more timely and appropriate care. However, this assumption is not supported by our results; in the intervention condition (BITSEA) there were relatively fewer referrals, but better psychosocial health at follow-up, compare to the control condition (KIPPPI).

We recommend future studies to evaluate the relationships between early detection, number and appropriateness of referrals, appropriateness of treatment, and health at follow-up.

Noteworthy is the fact that children in the control condition had a significantly higher CBCL Total Problem score at follow-up compared to baseline, despite the fact that more children in the control condition were referred. In the intervention condition there was no difference

between baseline and follow-up CBCL Total Problem score. So it seems to be that the impact of the BITSEA is prevention of deterioration in psychosocial problems in 2-years old. In the control condition, the instrument that is used is not scored by the child health professional, so possibly therefore its specificity is low, which might account for the more referrals.

We performed post-hoc analyses to evaluate whether the referred children were different between conditions with regard to the following variables: CBCL Total Problem score $M_{intervention}$= 25.5, $SD_{intervention}$=18.2; $M_{control}$=28.4, $SD_{control}$=19.7; Age $M_{intervention}$=23.80, $SD_{intervention}$=0.73; $M_{control}$=24.1, $SD_{control}$=1.17; Gender (boys) intervention: 48.2%; control: 58.9%; Ethnicity (native); intervention: 60.7%; control: 75.8%. So the referred children differed mainly in gender and ethnicity. We corrected our analyses for the potential confounding effects of these variables, so these differences between conditions may not be a conclusive explanation for the intervention effect. In order to further explore a possible explanation for the result that children in the intervention group had fewer psychosocial problems at follow-up, compared to children in the control condition, while children in the control condition were more often referred (7.9%) compared to children in the intervention condition (5.7%), we conducted additional analyses. The number of referrals for each condition was evaluated separately for children who had an 'at risk' score or a low score at baseline for the CBCL (>60); BITSEA Problem scale (score≥15 boys / ≥14 girls); and BITSEA Competence scale (score≤14 boys / ≤15 girls). We found no significant differences in number of referrals between conditions (see Table 7.5). The percentage of referrals in children with an 'at risk' score in the intervention condition, and the percentage of referrals in children with an 'at risk' score in the control condition do not differ significantly ($p>0.05$). We found an overall difference in number of referrals between conditions, but not when evaluated separately for 'at risk' and 'low scores', this might be explained by the relatively small size of the subgroups in these analyses.

Acceptability

Even though the BITSEA was scored, and the KIPPPI not, it took less time for parents and child health professionals to work with the BITSEA. Parents felt better prepared for the well-child visit after completing the KIPPPI, compared to the BITSEA. This might be, because the items of the KIPPPI are very structured and clustered per 'domain' (e.g. sleeping, eating, behaviour), whereas in the BITSEA items of the Problem and Competence scale are randomly arranged. This random arrangement might however contribute to less socially desirable answers, and therefore better validity.

Limitations and strengths

Our study has several limitations. First, we recognize that there are different ways in which psychosocial problems can be measured. We measured the primary outcome, psychosocial problems, with the CBCL. As stated in the introduction, the CBCL is a well-validated questionnaire, however it is a parent-completed instrument, rather than direct assessment. The report by parents introduces the proxy-problem; self-report by two-year-old children on their psychosocial problems is not possible, because children of this age lack the necessary language skills and the cognitive abilities to interpret the questions and they do not have a long-term view of events [47]. Therefore, proxy by parents may be a useful alternative [48].

Direct assessment of psychosocial problems, might have been an alternative for parent reported measurements. Direct assessment, for example by a psychologist, of every child in this study was not feasible with our large sample size. Furthermore, our data on the number of referrals was unable to provide a conclusive explanation for the difference between conditions in psychosocial problems at one year follow-up. Although child health professionals were instructed not to look at the CBCL answers, 11.4% of the child health professionals indicated that sometimes they did. The child health professional did not have the CBCL score, so referral decision could not be influenced by the CBCL score, but the answers might have influenced their assessment of the child. More elaborate data on referral decision could provide more insight in this underlying mechanism.

Second, as reported in the publication of the study protocol [33], at follow-up we measured health related quality of life (i.e. General Health Perceptions subscale and the Growth and Development subscale of the Infant Toddler Quality of Life questionnaire [ITQOL]) [39,49]. This variable was not measured at baseline, due to restrictions on the length of the questionnaire booklet at baseline. The child health centers were randomly allocated to the intervention and control condition, therefore similar health related quality of life in the conditions at baseline may be assumed. Therefore, despite the absence of longitudinal data on the ITQOL we evaluated the difference in health related quality of life between the control and intervention condition. We analyzed the difference between conditions uncorrected, as well as corrected for CBCL Total Problem score at baseline, and additionally for CBCL Total Problem score, gender and ethnicity of the child. We found no differences between conditions (data not shown).

Third, parents evaluated not the acceptability of the BITSEA and KIPPPI specifically, but the entire child health monitor questionnaire. However, the only differences between these

booklets, was the inclusion of either the BITSEA or the KIPPPI, so any differences between conditions may be attributed to these questionnaires.

Fourth, there were differences between intervention and control condition in drop-outs (respectively 59.2% and 68.9%). However, there was no difference between conditions in characteristics of people who were and were not retained in the study. So drop-out seemed to be non-selective and is not likely to have influenced the outcomes of the study. No information is available on the non-response group (i.e. parents that did not attend the well-child visit. It might be possible that parents avoid attending the well-child visit because they are afraid of possible interventions from Youth Care, but it might also be possible that parents do not find it necessary to attend the well-child visit because they feel confident that their child has no problems.

Fifth, the size and other characteristics of the clusters (i.e. child health care centers) varied in both the intervention and the control condition. Overall, differences between the intervention and control group were relatively small. But the differences between the clusters, especially the size (i.e. number of children in each cluster), may have lowered the power of the study. However, the power of this study was relatively large. No data, on cluster level, was available on deprivation indices, so we were unable to evaluate whether an unequal distribution between intervention and control clusters, with regard to the level of deprivation in the catchment area, had impact on the results of this study.

Lastly, the group of child health professionals who evaluated the acceptability of the BITSEA was rather small (n=105), however the response rate was very high (i.e. 92.2% of all child health professionals). The results reflect the situation in Rotterdam well, however generalizations to other areas or situations should be done with care.

Our study also has many strengths. First, our study design; random allocation of the well-child centers to the intervention and control condition eliminated systematic differences between these groups, which adds to the power of the study. Moreover, our longitudinal data made it possible to evaluate the impact of the early detection instrument under study over a longer period of time. Second, a major strength is that analyses were performed on a large and diverse community sample, which adds to the power of the results. Third, the questionnaires were applied in preventive child health care where they were used for the early detection of psychosocial problems. As the results suggests, it might be possible that child health care professionals do not always act solely on the BITSEA cutpoints if their professional expertise made them decide otherwise. See also the BITSEA manual [35], in which is stated that the

following things should also be taken into account while assessing the development of the child: level of worry of the parents; frequency or intensity of the behaviour; timing or onset of the behaviour; context of the behavior and the meaning of the behaviour in the cultural context of the parent. So the results reflect the outcomes as they will be in implemented, not just in theory.

Future research

We recommend that future studies evaluate more elaborately which children are being referred and follow these children more closely over a longer period of time. Additionally, the application of cut-points by the child health professional should be evaluated in more detail: to what extent are other factors (such as parental worry, situational or cultural context), important in the decision to refer?

This study was embedded within the Dutch system of preventive child health care provided by well-child centers in Rotterdam, the Netherlands. This may have consequences for the generalizability of our results in other areas and countries. Therefore, replication of our study in other, varied populations is recommended to confirm or reject the results presented in this study and to further explore pathways that explain the results of the present study.

Conclusion

The results support the use of the BITSEA as a tool for child health professionals in the early detection of psychosocial problem in 2-year-olds. It appeared that children experience fewer psychosocial problems when the BITSEA was used as compared to care as usual at one-year follow-up. The difference in psychosocial problems between BITSEA and 'care as usual' group at follow-up could not be explained by a difference in number of referrals. We recommend future studies in large and varied populations to replicate these findings.

REFERENCES

1. Briggs-Gowan MJ, Carter AS, Skuban EM, Horwitz SM (2001) Prevalence of social-emotional and behavioral problems in a community sample of 1-and 2-year-old children. Journal of the American Academy of Child and Adolescent Psychiatry 40: 811-819.
2. Klein Velderman M, Crone MR, Wiefferink CH, Reijneveld SA (2010) Identification and management of psychosocial problems among toddlers by preventive child health care professionals. European Journal of Public Health 20: 332-338.
3. Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP (2004) Identification and management of psychosocial problems among toddlers in Dutch preventive child health care. Archives of Pediatrics & Adolescent Medicine 158: 811-817.
4. Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP (2001) Identification and management of psychosocial problems by preventive child health care. Archives of Pediatrics & Adolescent Medicine 155: 462-469.
5. Zeijl E CM, Wiefferink K, Keuzenkamp S, Reijneveld M (SCP/TNO Kwaliteit van Leven: Den Haag/Leiden; 2005.) Kinderen in Nederland [Children in the Netherlands].
6. Durlak JA, Wells AM (1998) Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. American Journal of Community Psychology 26: 775-802.
7. Elliot J, Prior M, Merrigan C, Ballinger K (2002) Evaluation of a community intervention programme for preschool behaviour problems. Journal of Paediatrics and Child Health 38: 41-50.
8. Sturner RA (1991) Parent questionnaires - Basic office equipment. Journal of Developmental and Behavioral Pediatrics 12: 51-54.
9. Glascoe FP (2005) Screening for developmental and behavioral problems. Mental Retardation and Developmental Disabilities Research Reviews 11: 173-179.
10. Jellinek MS, Murphy JM (1999) Screening for psychosocial disorders in pediatric practice. American Journal of Diseases of Children 142: 1153-1157.
11. Briggs-Gowan MJ, Carter AS, Bosson-Heenan J, Guyer AE, Horwitz SM (2006) Are infant-toddler social-emotional and behavioral problems transient? Journal of the American Academy of Child and Adolescent Psychiatry 45: 849-858.
12. Egger HL, Angold A (2006) Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. Journal of Child Psychology and Psychiatry 47: 313-337.
13. Scheeringa MS (2001) The differential diagnosis of impaired reciprocal social interaction in children: A review of disorders. Child Psychiatry and Human Development 32: 71-89.
14. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington: VT: University of Vermont, Research Center for Children, Youth, and Families.

15. Briggs-Gowan MJ, Carter AS (1998) Preliminary acceptability and psychometrics of the Infant-Toddler Social and Emotional Assessment (ITSEA): A new adult-report questionnaire. Infant Mental Health Journal 19: 422-445.

16. Carter AS, Briggs-Gowan MJ, Jones SM, Little TD (2003) The Infant-Toddler Social and Emotional Assessment (ITSEA): Factor structure, reliability, and validity. Journal of Abnormal Child Psychology 31: 495-514.

17. Glascoe FP (2000) Early detection of developmental and behavioral problems. Pediatrics in Review 21: 272-280.

18. Eyberg S (1980) The Eyberg Child Behavior Inventory. Journal of Clinical Child Psychology 9: 22-28.

19. Mouton-Simien P, McCain AP, Kelley ML (1997) The development of the toddler behavior screening inventory. Journal of Abnormal Child Psychology 25: 59-64.

20. Bornstein MH, Hahn C-S, Haynes OM (2010) Social competence, externalizing, and internalizing behavioral adjustment from early childhood through early adolescence: Developmental cascades. Development and Psychopathology 22: 717-735.

21. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.

22. Kruizinga I, Jansen W, de Haan CL, van der Ende J, Carter AS, et al. (2012) Reliability and validity of the dutch version of the brief infant-toddler social and emotional assessment (BITSEA). PLoS ONE 7.

23. Nunnally JC, Bernstein IH (1994) Psychometric theory. New York: McGraw-Hill.

24. Terwee CB, Bot SDM, de Boer MR, van der Windt D, Knol DL, et al. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology 60: 34-42.

25. Kruizinga I, Jansen W, Mieloo CL, Carter AS, Raat H (2013) Screening accuracy and clinical application of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). PLoS ONE 8(8): e72602. doi:10.1371/journal.pone.0072602.

26. Kruizinga I, Visser JC, van Batenburg-Eddes T, Carter AS, Jansen W, et al. (2014) Screening for autism spectrum disorders with the brief infant-toddler social and emotional assessment. PLoS One 9: e97630.

27. Haapsamo H, Ebeling H, Soini H, Joskitt L, Larinen K, et al. (2009) Screening infants with social and emotional problems: A pilot study of the Brief Infant Toddler Social and Emotional Assessment (BITSEA) in Northern Finland. International Journal of Circumpolar Health 68: 386-393.

28. Karabekiroglu K, Briggs-Gowan MJ, Carter AS, Rodopman-Arman A, Akbas S (2010) The clinical validity and reliability of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). Infant Behavior & Development 33: 503-509.

29. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, et al. (2009) The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behavior & Development 32: 291-297.

30. Carter AS, Briggs-Gowan MJ, David NO (2004) Assessment of young children's social-emotional development and psychopathology: recent advances and recommendations for practice. Journal of Child Psychology and Psychiatry 45: 109-134.

31. Wendland J, Danet M, Gacoin E, Didane N, Bodeau N, et al. (2014) French Version of the Brief Infant-Toddler Social and Emotional Assessment Questionnaire-BITSEA. J Pediatr Psychol.

32. Kousemaker NPJ (1996) Zoeken, vinden, zorgen delen: de ontwikkeling van een praktijkparadigma voor onderkenning en pedagogische preventie van psychosociale problematiek in de Jeugdgezondheidszorg [Searching, finding, sharing care; the development of a practise paradigm for the descernment and pedagogical prevention of psychosocial problems in Youth Healthcare].

33. Kruizinga I, Jansen W, Carter AS, Raat H (2011) Evaluation of an early detection tool for social-emotional and behavioral problems in toddlers: The Brief Infant Toddler Social and Emotional Assessment - A cluster randomized trial. Bmc Public Health 11.

34. Hermans J, Ory F, Schrijvers G (Julius Centrum: Utrecht; 2005) Supporting development and parenting: sooner, faster and better: an advice for early detection and interventions in regard to developmental and parenting problems.

35. Briggs-Gowan MJ, Carter AS (2002) Brief Infant-Toddler Social and Emotional Assessment (BITSEA) mannual, version 2.0. New Haven: CT:Yale University.

36. Kruizinga I, Jansen W, de Haan CL, Raat H (2012) Reliability and Validity of the KIPPPI: An Early Detection Tool for Psychosocial Problems in Toddlers. PLoS ONE 7.

37. de Wolff MS, Theunissen MH, Vogels AG, Reijneveld SA (2013) Three questionnaires to detect psychosocial problems in toddlers: a comparison of the BITSEA, ASQ:SE, and KIPPPI. Academic pediatrics 13: 587-592.

38. Godoy L, Carter AS (2013) Identifying and Addressing Mental Health Risks and Problems in Primary Care Pediatric Settings: A Model to Promote Developmental and Cultural Competence. American Journal of Orthopsychiatry 83: 73-88.

39. Raat H, Landgraf JM, Oostenbrink R, Moll HA, Essink-Bot ML (2007) Reliability and validity of the Infant and Toddler Quality of Life Questionnaire (ITQOL) in a general population and respiratory disease sample. Quality of Life Research 16: 445-460.

40. Definitions 'native' and 'immigrant' (2012) - http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?conceptid=37. [Statistics Netherlands].

41. Cohen J (1988) Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erblaum Associates.

42. Cramér H (1999) Mathematical Methods of Statistics: Princeton Universtiy Press.

43. Kraemer S (2000) The fragile male. British Medical Journal 321:1609.

44. Reijneveld S, Harland P, Brugman E, Verhulst F, Verloove-Vanhorick S (2005) Psychosocial problems among immigrant and non-immigrant children - Ethnicity plays a role in their occurrence and identification. European Child and Adolescent Psychiatry 14: 145-152.

45. Vollebergh WAM, ten Have M, Dekovic M, Oosterwegel A, Pels T, et al. (2005) Mental health in immigrant children in the Netherlands. Social Psychiatry and Psychiatric Epidemiology 40: 489-496.

46. Killip S, Mahfoud Z, Pearce K (2004) What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. Annals of Family Medicine 2: 204-208.

47. Vogels T, Verrips GHW, Verloove-Vanhorick SP, Fekkes M, Kamphuis RP, et al. (1998) Measuring health-related quality of life in children: the development of the TACQOL parent form. Quality of Life Research 7: 457-465.

48. Theunissen NCM, Vogels TGC, Koopman HM, Verrips GHW, Zwinderman KAH, et al. (1998) The proxy problem: child report versus parent report in health-related quality of life research. Quality of Life Research 7: 387-397.

49. Raat H, Monhangoo AD, Grootenhuis MA (2006) Pediatric health-related quality of life questionnairese in clinical trials. Current Opinion in Allergy and Clinical Immunology 6: 180-185.

Chapter 8

General discussion

This thesis describes studies evaluating psychometric properties of the BITSEA in a Dutch general population sample of 2-year old children and whether health outcomes (i.e. psychosocial problems) were better when the BITSEA was applied, compared to care as usual; the KIPPPI. These aspects were evaluated for the total population as well as for subgroups by child gender and ethnicity.

The following research questions were addressed:

1. What is the reliability and validity of the BITSEA? (chapter 2)
2. What is the reliability and validity of the KIPPPI? (chapter 3)
3. What are the screening accuracy and cutpoints of the BITSEA? (chapter 4&5)
4. Is there a difference, a year after the well-child visit, in psychosocial wellbeing between children whose parents completed the BITSEA and children who received care as usual? (chapter 7)

In this chapter, the results of the studies presented in the thesis are discussed and interpreted. Strenghts and weaknesses in the methodology are addressed, and recommendations for future research and implications for public health practice are presented. At the end of the chapter an overall conclusion is given.

MAIN FINDINGS

What is the reliability and validity of the BITSEA?

In chapter 2 we described the reliability and validity of the BITSEA Problem and Competence scales. More specifically; internal consistency, test-retest reliability, parent-child care provider interrater reliability, concurrent validity (i.e. comparison to the Child Behavior Checklist [CBCL]) and discriminative validity were evaluated for the total population as well as for subgroups of child gender and ethnicity. The main finding was that the BITSEA Problem scale showed acceptable performance on all psychometric properties (*internal consistency* α=0.76; *test-retest reliability* ICC=0.75; *interrater reliability* ICC=0.30; *concurrent validity* Pearson coefficients ranging from 0.66 to 0.75; *discriminative validity* less favourable mean score for children over whom their parents worry, compared to children over whom their parent do not worry), whereas the BITSEA Competence scale showed acceptable performance on concurrent and discriminative validity (*internal consistency* α=0.63; *test-retest reliability* ICC=0.61; *interrater reliability* ICC=0.17; *concurrent validity* Pearson coefficients ranging from

-0.23 to -0.26; *discriminative validity* less favourable mean score for children over whom their parents worry, compared to children over whom their parents do not worry). There were no differences in the psychometric properties of the BITSEA between boys and girls or between native and non-native children.

The psychometric properties in this study are largely in line with what was found in previous studies on the BITSEA [1-5]. Compared to our results, other studies found somewhat higher internal consistency (Problem scale α=0.81; Competence scale α=0.72) [4], test-retest reliability (Problem scale ICC=0.87; Competence scale ICC=0.85) and parent-child care provider interrater reliability for the Competence scale (ICC=0.59) [1]. Contrary to our results, one study did not find a significant correlation between the BITSEA Competence scale and the CBCL Internalizing, Externalizing scores en Total Problem score [5]. Other studies did not evaluate the psychometric properties for subgroups. Differences in psychometric properties of the BITSEA between studies may be explained by difference in social demographic characteristics, sample size (ranging from 50 participants [3] to 3127 in our study), difference in setting (e.g. in the other studies the BITSEA was not used by a child health professional to assess the child's psychosocial development) and the use of a different version of the BITSEA [2].

What is the reliability and validity of the KIPPPI?

The reliability and validity of the KIPPPI is described in chapter 3. We evaluated the internal consistency, test-retest reliability, concurrent validity (i.e. comparison to CBCL), and discriminant validity. Additionally we evaluated the screening accuracy of the KIPPPI, relative to the CBCL; area under the receiver operating characteristic curve (AUC under the ROC) and for a range of KIPPPI scores the sensitivity, specificity. The main finding is that the psychometric properties of the KIPPPI Total score are adequate (*internal consistency* α=0.88; *test-retest reliability* ICC=0.80; *concurrent validity* Pearson coefficients ranging from 0.60-0.68; *discriminative validity* children over whom parents worried had less favourable scores compared to children over whom parents did not worry B=11.87, effect size=0.81); screening accuracy: AUC=0.85, sensitivity and specificity for the score with the highest Youden's index (i.e. score=54) respectively 0.87 and 0.81). The scales Wellbeing and Competence had comparable psychometric properties; however the scale Autonomy had inadequate psychometric properties.

Our results are largely in line with what is found in another study on the KIPPPI [2]. In that study, reliability, validity and screening accuracy of the KIPPPI was evaluated with a sample of 2106 2-year-old children. In that study, parents completed the KIPPPI and CBCL before coming to the well-child visit. The questionnaires were subsequently forwarded to the researchers and child health professionals were not informed about the results of the questionnaires. Children of immigrant origin were underrepresented. The internal consistency is equal between studies. However, we found a better sensitivity; the AUC in the other study was 0.85, by a pre-set specificity of 0.91, the sensitivity was 0.53. The authors do not mention the cutpoint that is associated with a specificity of 0.91 and a sensitivity of 0.53, so therefore we cannot compare that with our own results. The number of participants and the mean child age in both studies are comparable. The difference in sensitivity might be attributed to the setting in which the questionnaire was used; in our study the KIPPPI was used by the child health professional to assess de development of the child, whereas in the other study an anonymous research setting was chosen [2]. Parents might respond differently to sensitive topics when their answers are anonymous [6,7].

What are the screening accuracy and cutpoints of the BITSEA?

The screening accuracy of the BITSEA Problem scale was evaluated in comparison to a group of children with a Total Problem score in the clinical range of the Child Behavior Checklist (CBCL) (chapter 4) and both the Problem and Competence scale of the BITSEA were evaluated in comparison to a group of children with a diagnosis of autism spectrum disorder (ASD) (chapter 5). The AUC was calculated, and for a range of BITSEA scores the sensitivity, specificity, likelihood ratios (LHR), diagnostic odds ratio (OR) and Youden's index were calculated. The highest Youden's index was used to identify the (mathematical) optimal cutpoint.

Screening accuracy in comparison to a clinical CBCL score

We found a good screening accuracy [8] of the BITSEA Problem score when compared to a group of children with a CBCL Total Problem score in the clinical range, AUC=0.97 (95%CI=0.95-0.98). The highest Youden's index (i.e. 0.85) was found for Problem score 14. The associated indices for screening accuracy were sensitivity=0.95; specificity=0.90; LHR+=9.38; LHR-=0.05; OR=181.20. There were no differences in screening accuracy between boys and girls nor between native and non-native children (i.e. 95%CI of the AUC

did not overlap). In our study, the same optimal cutpoint was found for boys and girls. We found different optimal cutpoints for native and non-native children, where native children differed from the other (sub)samples in cuptoint as indicated by the Youden index; score 17. However the screening accuracy between native and non-native children was equal. The application of different cutpoints for different ethnic groups in preventive child health care may not be desirable, since it is difficult to determine whether the different distribution and mean BITSEA scores can be attributed to the actual amount or seriousness of problems, or that it reflects cultural differences (e.g. interpretation of behaviour, or question items). Moreover, the compositions of ethnic groups may change over time, which would mean the (continuous) evaluation and adjustment of cutpoints.

Screening accuracy in comparison to a diagnosis of ASD

We found a good screening accuracy [8] of the BITSEA Problem and Competence score and the newly calculated Autism score when compared to a group of children with a diagnosis of ASD, respectively, AUC=0.90 (95%CI=0.87-0.92); 0.93 (95%CI=0.91-0.95); 0.95 (95%CI=0.93-0.97). The highest Youden's index (i.e. respectively J=0.67; J=0.74; J=0.78) was found for Problem score 13; Competence score 15; Autism score 10. The associated indices for screening accuracy was, for the Problem scale; sensitivity=0.83; specificity=0.84; LHR+=5.18; LHR-=0.20; OR=26.22. For the Competence scale: sensitivity =0.85; specificity=0.89; LHR+=7.92; LHR-=0.17; OR=47.95. And for the Autism score; sensitivity=0.85; specificity=0.93; LHR+=12.40; LHR-=0.16; OR=78.79. The screening accuracy of the newly calculated Autism score was for the total population better than the screening accuracy of the Problem score, but not better than the Competence score. The Autism score if formally not a separate BITSEA scale and the findings of this study imply that calculation of the Autism score is unnecessary when the Competence score is known. The Youden index yielded the same cutpoints for boys and girls for the Problem (i.e. score 13) and Competence scale (i.e. score 15). The optimal cutpoint for the Autism score for the total population was 10, for boys it was 9 and for girls 8.

The screening accuracy in our study was comparable with that found in other studies [1,2]. We determined the screening accuracy for children aged 24 months, so also the cutpoints are determined for this age. It is expected that children with typical development acquire more competencies with age, previous studies have found higher Competence scores in older children, compared to younger children [1,9]. In the United States separate cutpoints were set

by gender for the Problem scale because the distributions differed at the extremes, such that a universal cutpoint would identify unequal proportions across age by sex groups. Cutpoints for Problems and Competence scores were defined in 6-month age groups by child's sex. The sample in our study was quite homogeneous regarding age (M=23.7 months, SD=0.7), so we cannot generalize our results to other age groups. At follow-up we could evaluate the screening accuracy of the BITSEA Problem scale again for child aged 36 months, with the CBCL Total Problem score as a criterion. The BITSEA Competence scale could not be evaluated because the content of the BITSEA Competence items do not resemble the content of the items on the CBCL Total Problem score. We found an equal screening accuracy compared to children aged 24 months. The optimal cutpoint at age 36 months is score 11, which is somewhat lower than the optimal cutpoint we found at age 24 (i.e. score 14) and what was found in another study of children aged 30-35 months (girls score 12 and boys score 14) [9].

Is there a difference, a year after the well-child visit, in psychosocial wellbeing between children whose parents completed the BITSEA and children who received care as usual?

Children in the intervention condition scored more favourably on the CBCL at one year follow-up than children in the control condition: B=-2.43 (95% confidence interval [95%CI] = -3.53;-1.33 $p<0.001$). Child health professionals reported to refer fewer children in the intervention condition (n=56, 5.7%), compared to the control condition (n=95, 7.9%; $p<0.05$). Noteworthy is the fact that children in the control condition had a significantly higher CBCL Total Problem score at follow-up compared to baseline. In the intervention condition there was no difference between baseline and follow-up CBCL Total Problem score. So it seems to be that the impact of the BITSEA in this study was the prevention of deterioration in psychosocial problems in 2-years old. The percentage of referrals in children with an 'at risk' score in the intervention condition, and the percentage of referrals in children with an 'at risk' score in the control condition did not differ significantly.

METHODOLOGICAL CONSIDERATIONS

Study design

In the studies we used different study designs; cross-sectional (chapters 2 to 5) and longitudinal (i.e. cluster randomized controlled design in chapter 7). In a cross sectional study design

data is collected at one specific point in time. This type of studies is descriptive in nature and is often used to identifying associations or to determine prevalence [10]. In the present study the cross-sectional study design was used to evaluate psychometric properties of questionnaires. In a cluster randomized controlled trial (RCT), groups of individuals, rather than individuals, are randomly assigned to either the intervention condition or the control condition. The participants in such a study cannot be regarded as statistically independent, therefore cluster randomization designs tend to be less efficient than designs which randomize individuals, however it is necessary when randomization on individual level is not feasible due to practical considerations [11]. It is quite unique that a study on the effect of an early detection questionnaire is conducted with a RCT. However, the completion of the BITSEA by parents is not seen as an intervention in itself. The work processes of the child health professionals to assess the psychosocial development of children differed between conditions. Also, the professionals in the intervention condition received additional training in, among others, the early detection of psychosocial problems (whereas the professionals in the control condition did not). The effect of the BITSEA as an early detection tool has not been evaluated with this method before.

Studies performed in 'daily practice' often have a quasi-experimental study design [12]. Studies with a quasi-experimental design lack the element of (correctly performed) random assignment to treatment or control condition. The main advantage of a (cluster) randomized controlled trial is the stronger internal validity, because, due to random assignment, the intervention and control condition may be assumed to be comparable at baseline. With random assignment, participants (or clusters) have the same chance of being assigned to the intervention or control condition. Therefore, differences between groups at baseline would be due to chance, rather than to a systematic factor related to the intervention [13].

It was not possible to conduct a double blind study; both parents and child health care providers were aware of the type of questionnaire that was used. Some child health professionals may have been aware that the BITSEA was recommended as 'a promising questionnaire' by an advisory committee to the Dutch government [14], which could have influenced their assessment of the child and subsequent (referral) decisions.

When health professionals use the BITSEA, they also use the cutpoints for scores that indicate the possible presence or absence of problems and delays in competencies. This may help the professionals to provide parents with adequate advice with regard to the presence of problems and the need for referral. In the care as usual setting in Rotterdam,

the KIPPPI is predominantly used as a tool to structure discussion between parent and child health professional, and less as a screening instrument in a formal way, given the absence of empirically determined cutpoints. In the current study, with our data, we were not able to determine the mechanisms behind the findings that children experience fewer psychosocial problems when the BITSEA was applied as compared to care as usual at one year follow-up. Therefore we must be careful with the interpretation of these findings, and we recommend future studies to replicate these findings.

Control condition

The intervention condition was being evaluated against 'care as usual' as a control condition. Care as usual consisted of administering the KIPPPI. Many aspects of psychosocial problems are addressed in the KIPPPI; therefore this questionnaire can be used by the child health professional to guide conversation with the parent. Based on the conversation with the parent and the completed KIPPPI, the child health professional made a clinical decision whether a child is to be referred to a mental health care professional or whether a follow-up consultation is required. Although the KIPPPI can be scored, in this study the child health care professionals did not calculate scores since cutpoints were not empirically determined at the time. It is generally acknowledged that the use of questionnaires facilitate the early detection of psychosocial problems [2,15-20], moreover, using an instrument for the early detection of psychosocial problems is strongly recommended by the Health Care Inspection of the Dutch Government [21,22]. Therefore it was not an ethical option to have a control condition with well-child visits where no early detection was used. So the effectiveness of the BITSEA was not evaluated in an absolute way, but relative to another early detection instrument. As a consequence we did not expect to find a large effect of condition on psychosocial problems at one year follow-up.

Study population and response

A characteristic of the studies in this thesis is that they were carried out in the setting of well-child clinics. The well-child clinics are the main points of access for preventive child health care for children 0-4 years old. Although a main advantage of the well-child clinic setting is the relatively high response rate of completed questionnaires from the parents that attended the well-child visit (61.4%), people not attending the well-child clinic may have led to selection bias. Therefore, results should be interpreted with caution. For example, parents may not attend the well-child visit, because they might be afraid of possible interventions

from Youth Care, but it might also be possible that parents do not find it necessary to attend the well-child visit because they feel confident that their child has no problems.

A strength of the study is the large and diverse sample size. The large and diverse sample of respondents made it possible for us to evaluate research questions in subsamples of child gender and ethnicity. For the evaluation of the screening accuracy of the BITSEA with regard to autism spectrum disorders (ASD), we included a sample of children diagnosed with ASD by experienced clinicians and diagnoses were based on extensive multidisciplinary diagnostic procedures.

Additionally, the practice-based setting inherent to preventive child health care can be considered a strength. The advantage of a practice-based approach is its basis in daily practice rather than in experimental or laboratory situations. This makes it unnecessary to do assumptions on how the intervention would work in practice. With that, it makes it easier to implement the new interventions.

Measurements

In the studies we evaluated questionnaires in which parents provided a self-reported account of their child's well-being. It is well known that responses to questions may not reflect a person's true status. People may distort their answers in order to appear more favourably. As a result, the tendency to present oneself (or in this case, their child) in the best possible light may systematically bias responses toward respondents' perceptions of what is socially acceptable [23]. Questions that are sensitive tend to produce relatively higher non-response rates or larger measurement error in responses than questions on other topics [23]. A question may be sensitive if it raises fears about the likelihood or consequences of disclosure of the answer to other agencies or individuals [23], such as the child health professional. Parents may provide too favourable answers, in order to avoid interventions from Youth Care. The opposite is also a plausible scenario; there might be reasons why parents may provide less favourable answers than the reality. Some studies show that depressed mothers tend to have a negative bias towards the perceptions of their child's behaviour and emotions. [24-27] From these studies it is unclear if depressed mothers' negative views of their children are accurate, or if they represent depression-induced cognitive distortions. Some studies suggest that the difference between depressed and non-depressed mothers' perceptions of their children is not based on the negative affective bias associated with depression but on the tendency of non-depressed mothers to be more positive about their children than is actually warranted (i.e. a

positive affective bias) [26]. However, in a review of 22 studies, Richters [28] challenges the empirical foundation for the widely held assumption that depressed mothers have distorted perceptions of their children's problems. In our study, we cannot report on the effect of parental depression on BITSEA scores, since we did not enquire psychosocial problems of the parents in our study.

Self-report by two-year-olds children on their psychosocial problems is not possible, because children of this age lack the necessary language skills and the cognitive abilities to interpret the questions and they do not have a long-term view of events [29]. Therefore, proxy by parents may be a useful alternative [30].

Since all questionnaires used for well-child visits are in Dutch, in our study the BITSEA was also only available in Dutch. This may also result in selection bias. Non-native people who do not speak the language well may have trouble understanding the questionnaires, answering, or may not even try. People were encouraged to ask for help by relatives for completing the questionnaire, also at the well-child clinic help could be provided. To evaluate the quality of the given answers, we asked parents whether they understood all questions well. Of all parents 77.1% indicated that they understood all questions well.

IMPLICATION FOR PRACTICE

This study adds to the knowledge on the use of short comprehensive instruments that are appropriate to measure psychosocial problems in children of preschool age. Existing instruments, such as the Eyberg Child Behavior Inventory [31] or the Toddler Behavior Screening Inventory [32], only measure problem behaviour and do not address social-emotional competencies. Measuring delays in social-emotional competence, however, is also important since delays in competence are for instance related to internalizing and externalizing problems later in life [33]. There remains a need for a short instrument that measures both problems and delays in competence. This study showed that another early detection instrument, the BITSEA, has acceptable psychometric properties and that it is useful in the preventive child health care. This enlarges the selection of instruments to measure psychosocial problems in young children.

Other benefits of the BITSEA are that is a relatively short instrument (i.e. 42 items) and its international use. International prevalence rates can be compared, which might provide relevant input for policy.

The acceptability of the BITSEA as perceived by parents and professionals was generally good. It took parents and professionals less time to complete and administer, respectively, the BITSEA in comparison to care as usual (i.e. KIPPPI). Parents in the control condition, however, felt even better prepared for the well-child visit, compared to the intervention condition. The BITSEA has adequate and equal psychometric properties for subgroups of child gender and ethnicity, so the BITSEA can be used for the assessment of psychosocial problems in a heterogeneous population.

FUTURE RESEARCH

Our data on referral decisions as registered by the child health professional at baseline did not enable us to explain this effect of the BITSEA on the fewer psychosocial problems a year after the well-child visit, so the mechanisms behind this difference are still unknown. We recommend that our study will be replicated in other populations. Processes regarding quality of referrals and received care should be evaluated more elaborately. In future studies we need more information on referral processes; which children exactly are being referred, to what kind of mental health professional, within what timeframe, with a more reliable measurement of pursuit of referrals.

The BITSEA is intended for the early detection of psychosocial problems in 1-3 years olds. In our study we evaluated the BITSEA for children aged 2 years old. Findings on sensitivity and specificity cannot merely be generalized to other age group, which means that other cutpoints should be applied for older and younger children. For the use of the BITSEA in a broader age spectrum, additional research is needed to identify appropriate cutpoints for other ages.

We evaluated the screening accuracy of the BITSEA regarding a clinical Total Problem score and regarding an ASD-diagnosis. Future studies may further evaluate the screening accuracy of both the BITSEA Problem and Competence scale with reference groups of children with a broad range of psychosocial problems, or in homogeneous groups of children with specific psychosocial or psychiatric problems, who are diagnosed by a mental health professional.

We followed the two-year-old children over a period of one year. We measured whether the children were referred in that year and whether that referral was pursued. The formal diagnoses made at the earliest age often is ASD; on average the first symptoms to arouse parental concerns about children eventually diagnosed with ASD occur before the age of two years. However, the average age of ASD diagnosis is approximately three years of age and

often occurs later [34]. Therefore we recommend future studies to apply a longer follow-up period and to follow the children closely after a referral was made.

Because the time constraints of a well-child visit (i.e. 25 minutes are reserved for a well-child visit), we would recommend studies on a digital version of the BITSEA. Scoring of the BITSEA requires child health care professionals to add the answers given on the questionnaire, which cost them about 5 minutes. Implementing new work processes that allow for digital completion of the questionnaire for parents, and automized scoring, might save valuable time. The psychometric properties could be evaluated and compared to the psychometric properties of the paper-and-pencil version. Also the acceptability for both the child health care professionals and parents could be evaluated.

GENERAL CONCLUSION

The studies described in this thesis indicate that the BITSEA is a questionnaire with good psychometric properties. The Problem scale has better psychometric properties than the Competence scale, however, both scales discriminate children with and without problems well. The KIPPPI, that was used in the control condition (i.e. care as usual) also has good psychometric properties; however currently cutpoints to indicate children with possible problems are not used when applying the KIPPPI. The BITSEA has several strong points: The BITSEA is a relatively short questionnaire which takes only little time to complete and administer. The BITSEA can be scored and cutpoints, that indicate a risk on problem behaviour and/or delays in competencies and that can be applied in practice are available (they were determined in this thesis). Since the BITSEA is an internationally applied questionnaire, prevalence of psychosocial problems can be compared between countries. It appeared that children experience fewer psychosocial problems when the BITSEA was used as compared to care as usual at one-year follow-up; we recommend future studies to replicate these findings. This study supports the BITSEA as a useful instrument, for the use in preventive child health care in the Netherlands, to detect psychosocial problems at an early age and in an early stage.

REFERENCES

1. Velderman M, Crone M, Wiefferink C, Reijneveld S (2010) Identification and management of psychosocial problems among toddlers by preventive child health care professionals. European Journal of Public Health 20: 332-338.
2. Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP (2004) Identification and management of psychosocial problems among toddlers in Dutch preventive child health care. Archives of Pediatrics & Adolescent Medicine 158: 811-817.
3. Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP (2001) Identification and management of psychosocial problems by preventive child health care. Archives of Pediatrics & Adolescent Medicine 155: 462-469.
4. Briggs-Gowan MJ, Carter AS, Skuban EM, Horwitz SM (2001) Prevalence of social-emotional and behavioral problems in a community sample of 1-and 2-year-old children. Journal of the American Academy of Child and Adolescent Psychiatry 40: 811-819.
5. Verhulst FC, van der Ende J (1997) Factors associated with child mental health service use in the community. Journal of the American Academy of Child and Adolescent Psychiatry 36: 901-909.
6. Briggs-Gowan MJ, Carter AS (2008) Social-emotional screening status in early childhood predicts elementary school outcomes. Pediatrics 121: 957-962.
7. Campbell SB, Spieker S, Burchinal M, Poe MD (2006) Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. Journal of Child Psychology and Psychiatry 47: 791-800.
8. Hakman M, Sullivan M (2009) The Effect of Task and Maternal Verbosity on Compliance in Toddlers. Infant and Child Development 18: 195-205.
9. Mesman J, Koot HM (2001) Early predictors of preadolescent internalizing and externalizing DSM-IV diagnoses. J Am Acad Child Adolesc Psychiatry 40: 1029-1036.
10. Moffitt TE, Caspi A (2001) Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. Development and Psychopathology 13: 355-375.
11. American Academy of Pediatrics. Committee on children with disabilities: development surveillance and screening of infants and young children. (2001) Pediatrics 108: 192-196.
12. Bricker D, Schoen Davis M, Squires J (2004) Mental health screening in young children. Infants and Young Children 17: 129-144.
13. (2003) Horen, zien, niet zwijgen [See all, hear all, say something]. Inspectie voor de Gezondheidszorg, Inspectie van het Onderwijs, Inspectie Openbare Orde en Veiligheid, Inspectie Jeugdhulpverlening, Jeugdbescherming.
14. Durlak JA, Wells AM (1998) Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. American Journal of Community Psychology 26: 775-802.

15. Elliot J, Prior M, Merrigan C, Ballinger K (2002) Evaluation of a community intervention programme for preschool behaviour problems. Journal of Paediatrics and Child Health 38: 41-50.

16. Basistakenpakket Jeugdgezondheidszorg 0-19 jaar [Youth Health Care Tasks 0-19 years]. (2002). Ministerie van Volksgezondheid, Welzijn en Sport.

17. (2009) Centraal Bureau voor de Statistiek [Statistics Netherlands]. Statistics Netherlands.

18. Kousemaker NPJ (1996) Zoeken, vinden, zorgen delen: de ontwikkeling van een praktijkparadigma voor onderkenning en pedagogische preventie van psychosociale problematiek in de Jeugdgezondheidszorg [Searching, finding, sharing care; the development of a practise paradigm for the descernment and pedagogical prevention of psychosocial problems in Youth Healthcare].

19. Briggs-Gowan MJ, Carter AS, Irwin JR, Wachtel K, Cicchetti DV (2004) The brief infant-toddler social and emotional assessment: Screening for social-emotional problems and delays in competence. Journal of Pediatric Psychology 29: 143-155.

20. Hermans J, Ory F, Schrijvers G (Julius Centrum: Utrecht; 2005) Supporting development and parenting: sooner, faster and better: an advice for early detection and interventions in regard to developmental and parenting problems.

21. Achenbach TM, Edelbrock C, Howell CT (1987) Empirically based assessment of the behavioral/emotional problems of 2- and 3-year-old children. J Abnorm Child Psychol 15: 629-650.

22. Karabekiroglu K, Rodopman-Arman A, Ay P, Ozkesen M, Akbas S, et al. (2009) The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). Infant Behavior & Development 32: 291-297.

23. Haapsamo H, Ebeling H, Soini H, Joskitt L, Larinen K, et al. (2009) Screening infants with social and emotional problems: A pilot study of the Brief Infant Toddler Social and Emotional Assessment (BITSEA) in Northern Finland. International Journal of Circumpolar Health 68: 386-393.

24. Achenbach TM, McConaughy SH, Howell CT (1987) Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity Psychological Bulletin 101: 213-232.

25. Wild D, Grove A, M Martin, al. e (2005) Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 8: 94-104.

26. Briggs-Gowan MJ, Carter AS (2002) Brief Infant-Toddler Social and Emotional Assessment (BITSEA) mannual, version 2.0. New Haven: CT:Yale University.

27. Verhulst FC, Van der Ende J, Koot HM (1997) Handleiding voor de Teacher's Report Form (TRF). [Manual for the Theacher's Report Form (TRF)]. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

28. Raat H, Landgraf JM, Oostenbrink R, Moll HA, Essink-Bot ML (2007) Reliability and validity of the Infant and Toddler Quality of Life Questionnaire (ITQOL) in a general population and respiratory disease sample. Quality of Life Research 16: 445-460.

29. Raat H, Monhangoo AD, Grootenhuis MA (2006) Pediatric health-related quality of life questionnairese in clinical trials. Current Opinion in Allergy and Clinical Immunology 6: 180-185.

30. Campbell MK, Elbourne DR, Altman DG, Grp C (2004) CONSORT statement: extension to cluster randomised trials. British Medical Journal 328: 702-708.

31. Twisk JWR (2006) Applied Multilevel Analysis: A practical guide. Cambridge: Cambridge University Press.

32. Achenbach TM, Rescorla LA (2000) Manual for the ASEBA Preschool Forms & Profiles. Burlington: VT: University of Vermont, Research Center for Children, Youth, and Families.

Summary

The prevalence of psychosocial problems is already relatively high in preschool children. The prevalence rates differ between studies, depending on the measuring method that is used. In the Netherlands, 5-10% of the preschool children score in the clinical range on a questionnaire that measures psychosocial problems. Psychosocial problems is a collective term for a great variety of problems. Often these problems are divided in two components; externalizing and internalizing problems.

- Externalizing problems includes problems such as; aggressive behaviour or overactivity.
- Internalizing problems includes problems such as; anxiety, depression or withdrawal.

It has been recommended that psychosocial problems should be detected at a very young age and followed by appropriate care. Measurement, early detection and treatment of psychosocial problems at a young age are important because this may contribute to a reduction of problems and an increase of competencies at older ages. Early detection of psychosocial problems is more effective when parental completed questionnaires are used, such as the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). The central aim of this study is to evaluate the reliability and validity of the Dutch version of the BITSEA in 2-year old children (part one of this thesis). Also, in a cluster randomized trial, the aim was to assess whether health outcomes (specifically psychosocial well-being) were better in the intervention group where the BITSEA was applied to detect psychosocial problems compared to the outcomes in the control group where 'care as usual' was applied (i.e. use of the KIPPPI); the follow-up duration was one year (part two of this thesis).

The following research questions are addressed in this thesis:

Question 1: What is the reliability and validity of the BITSEA for the total population as well as for subgroups of child gender and ethnicity?

In chapter 2 we concluded that the BITSEA Problem scale showed acceptable performance on all psychometric properties (i.e. internal consistency, test-rest reliability, interrater reliability, concurrent validity and discriminative reliability). The BITSEA Competence scale showed acceptable performance on concurrent and discriminative validity. There were no differences in the psychometric properties between boys and girls or between native and non-native children.

Question 2: What is the reliability and validity of the KIPPPI?

In chapter 3 we concluded that the psychometric properties (i.e. internal consistency, test-retest reliability, concurrent validity, discriminative validity and screening accuracy) of the KIPPPI Total score are adequate. The scales Wellbeing and Competence had psychometric properties comparable to the Total score, however the scale Autonomy had inadequate psychometric properties.

Question 3: What are the screening accuracy and cutpoints of the BITSEA for the total population as well as for subgroups of child gender?

In chapter 4, the screening accuracy of the BITSEA Problem scale is evaluated in comparison to a group of children with a Total Problem score of the Child Behavior Checklist (CBCL) in the clinical range. In chapter 5, the screening accuracy of both the Problem and Competence scale of the BITSEA were evaluated in comparison to a group of children with a diagnosis of autism spectrum disorder (ASD). The area under the curve (AUC) was calculated, and for a range of BITSEA scores the sensitivity, specificity, likelihood ratio's (LHRs), diagnostic odds ratio (OR) and Youden's index were calculated. The highest Youden's index was used to identify the (mathematical) optimal cutpoint.

We found a good screening accuracy of the BITSEA Problem score when compared to a group of children with a CBCL Total Problem score in the clinical range. There were no differences in screening accuracy between boys and girls or between native and non-native children. In our study, the same optimal cutpoint was found for boys and girls. We found different optimal cutpoints for native and non-native children, where native children differed from the other (sub)samples. The application of different cutpoints for different ethnic groups in preventive child health care is however not advised, since it is difficult to determine whether the different distribution and mean BITSEA scores can be attributed to the actual amount or seriousness of problems, or that it reflects cultural differences (e.g. interpretation of behaviour, or question items). Moreover, the compositions of ethnic groups may change over time, which would mean the (continuous) evaluation and adjustment of cutpoints. The application of the optimal cutpoint identified for the total population will lead to an over referral in native children.

We found a good screening accuracy of the BITSEA Problem and Competence score and the newly calculated Autism score when compared to a group of children with a diagnosis of ASD. The screening accuracy of the newly calculated Autism score was for the total population better than the screening accuracy of the Problem score, but not better than

the Competence score. The Autism score is formally not a separate BITSEA scale and the findings of this study imply that calculation of the Autism score is unnecessary when the Competence score is known. The Youden index yielded the same cutpoints for boys and girls for the Problem and Competence scale.

Question 4: Is there a difference, a year after the well-child visit, in psychosocial wellbeing between children whose parents completed the BITSEA and children who received care as usual?

The study design for providing an answer to this question is described in chapter 6, the results are presented in chapter 7. Children in the intervention condition scored more favourably on the CBCL at one year follow-up, than children in the control condition. Child health professionals reported to refer fewer children in the intervention condition, compared to the control condition. The percentage of referrals in children with an 'at risk' score in the intervention condition, and the percentage of referrals in children with an 'at risk' score in the control condition did not differ significantly.

Finally, in chapter 8, the study findings were integrated and the main results for the research questions were summarized and discussed. Implications for further research and practice are presented.

The studies described in this thesis indicate that the BITSEA is a questionnaire with good psychometric properties. The Problem scale has better psychometric properties than the Competence scale, however, both scales discriminate children with and without problems well. The KIPPPI, that was used in the control condition (i.e. care as usual), also has good psychometric properties; however currently cutpoints to indicate children with possible problems are not used when applying the KIPPPI. The BITSEA has several strong points: The BITSEA is a relatively short questionnaire which takes only little time to complete and administer. The BITSEA can be scored and cutpoints, that indicate a risk on problem behaviour and/or delays in competencies and that can be applied in practice are available (they were determined in this thesis). Since the BITSEA is an internationally applied questionnaire, prevalence of psychosocial problems can be compared between countries.

It appeared that children experience fewer psychosocial problems when the BITSEA was used as compared to care as usual at one-year follow-up; we recommend future studies to replicate these findings. This study supports the BITSEA as a useful instrument, for the use in preventive child health care in the Netherlands, to detect psychosocial problems at an early age and in an early stage.

Samenvatting

De prevalentie van psychosociale problemen is al relatief hoog bij peuters. De prevalentiecijfers in onderzoeken verschillen van elkaar, dit komt doordat verschillende meetmethodes zijn gebruikt. In Nederland scoort 5-10% van de peuters in de klinische range op een vragenlijst die psychosociale problemen meet. 'Psychosociale problemen' is een verzamelnaam voor een verscheidenheid aan problemen. Deze problemen worden vaak opgedeeld in twee categorieën; externaliserende en internaliserende problemen.

- Externaliserende problemen omvatten problemen zoals agressief gedrag of hyperactiviteit.
- Internaliserende problemen omvatten problemen zoals angst, depressie of teruggetrokkenheid.

Er wordt geadviseerd om psychosociale problemen te signaleren op een jonge leeftijd en dit te laten opvolgen door passende zorg. Het meten, vroeg signaleren en behandelen van psychosociale problemen op jonge leeftijd is belangrijk omdat dit kan bijdragen aan een vermindering van problemen en een toename van competenties op latere leeftijd. Vroegsignalering van psychosociale problemen is effectiever wanneer vragenlijsten worden gebruikt die door ouders zijn ingevuld. Een voorbeeld van een dergelijke vragenlijst is de Brief Infant-Toddler Social and Emotional Assessment (BITSEA). Het doel van dit proefschrift was het evalueren van de betrouwbaarheid en validiteit van de Nederlandse versie van de BITSEA voor 2-jarige kinderen (deel één van dit proefschrift). Daarnaast is een clustergerandomiseerd onderzoek gedaan om te bepalen of gezondheidsuitkomsten (in het bijzonder psychosociaal welzijn) beter waren in de interventiegroep waar de BITSEA werd toegepast om psychosociale problemen te signaleren, in vergelijking met de uitkomsten in de controlegroep waar 'zorg zoals gebruikelijk' werd toegepast (namelijk het gebruik van de KIPPPI); de opvolgperiode was 1 jaar (deel twee van dit proefschrift).

De volgende onderzoeksvragen komen aan bod in dit proefschrift:

Vraag 1: Wat is de betrouwbaarheid en validiteit van de BITSEA voor de hele populatie en voor subgroepen voor wat betreft geslacht en etniciteit van het kind?

In hoofdstuk 2 concluderen we dat de BITSEA Probleemschaal acceptabel presteert op alle psychometrische eigenschappen (namelijk interne consistentie, test-hertestbetrouwbaarheid, interbeoordelaarsbetrouwbaarheid, concurrente validiteit en discriminante validiteit). The BITSEA Competentieschaal presteert acceptabel op concurrente validiteit en discriminante validiteit. Er waren geen verschillen in psychometrische eigenschappen tussen jongens en meisjes en tussen autochtone en allochtone kinderen.

Vraag 2: Wat is de betrouwbaarheid en validiteit van de KIPPPI?

In hoofdstuk 3 concluderen we dat de psychometrische eigenschappen (namelijk interne consistentie, test-hertestbetrouwbaarheid, concurrente validiteit, discriminante validiteit en nauwkeurigheid in het signaleren van problemen) van de KIPPPI Totaalscore adequaat zijn. De schalen Welzijn en Competentie hebben vergelijkbare psychometrische eigenschappen als de Totaalscore. De psychometrische eigenschappen van de schaal Autonomie zijn daarentegen inadequaat.

Vraag 3: Wat is de nauwkeurigheid bij het signaleren van psychosociale problemen en wat zijn de afkappunten van de BITSEA voor de totale populatie en voor subgroepen voor wat betreft geslacht van het kind.

In hoofdstuk 4 wordt de nauwkeurigheid bij het signaleren van psychosociale problemen van de BITSEA Probleemschaal geëvalueerd in vergelijking met de totale probleemscore van de Child Behavior Checklist (CBCL) in de klinische range. In hoofdstuk 5 wordt zowel de Probleemschaal als de Competentieschaal van de BITSEA geëvalueerd op onderscheidend vermogen in een algemene populatie in vergelijking met een groep kinderen met een autismespectrumstoornisdiagnose. De 'area under the curve' is berekend. De hoogste Youden's index is gebruikt om het (mathematisch) optimale afkappunt te bepalen.

We vonden een goede nauwkeurigheid bij het signaleren van psychosociale problemen met de BITSEA Probleemscore bij vergelijking met een groep kinderen met een CBCL totale probleemscore in de klinische range. Er waren geen verschillen wat dit betreft tussen jongens en meisjes en ook niet tussen autochtone en allochtone kinderen. In onze studie vonden we hetzelfde optimale afkappunt voor jongens en meisjes. We vonden verschillen in optimale afkappunten voor autochtone en allochtone kinderen. Autochtone kinderen verschilden van de andere (sub)groepen. Het toepassen van verschillende afkappunten voor verschillende etnische groepen in de preventieve jeugdgezondheidszorg wordt desondanks niet aanbevolen. Dit omdat het moeilijk vast te stellen is of de verschillende verdelingen en gemiddelde BITSEA-scores kunnen worden toegeschreven aan de daadwerkelijke hoeveelheid problemen, aan de ernst van de problemen, of dat het een reflectie is van culturele verschillen (bijvoorbeeld interpretatie van gedrag of de vragenlijstitems). Bovendien kan de samenstelling van etnische groepen veranderen en dit zou kunnen betekenen dat de afkappunten continu geëvalueerd en aangepast moeten worden. Als het optimale afkappunt voor de totale populatie wordt gebruikt, resulteert dat in oversignalering van problemen bij autochtone kinderen.

We vonden een goede nauwkeurigheid bij het signaleren van psychosociale problemen met de BITSEA Probleem- en Competentiescore bij een vergelijking met een groep kinderen met een autismespectrumstoornisdiagnose. De nauwkeurigheid bij het signaleren van psychosociale problemen met de nieuw berekende Autismescore was voor de totale populatie beter dan de nauwkeurigheid van signaleren met de Probleemscore, maar niet beter dan met de Competentiescore. De Autismescore is formeel geen aparte BITSEA-schaal en de resultaten van deze studie impliceren dat het berekenen van de Autismescore niet nodig is wanneer de Competentiescore bekend is. De Youdenindex gaf aan dat de afkappunten voor jongens en meisjes gelijk zijn voor de Probleem- en Competentieschaal.

Vraag 4: Is er, een jaar na het consult, een verschil in psychosociaal welzijn tussen kinderen van wie de ouders de BITSEA hebben ingevuld en kinderen die 'zorg zoals gebruikelijk' ontvingen?

De onderzoeksmethode om deze vraag te beantwoorden wordt beschreven in hoofdstuk 6, de resultaten worden gepresenteerd in hoofdstuk 7. Kinderen in de interventieconditie scoorden een jaar na het consult gunstiger op de CBCL dan kinderen in de controlegroep. Professionals in de jeugdgezondheidszorg rapporteerden dat ze minder kinderen hebben doorverwezen in de interventieconditie dan in de controleconditie. Het percentage doorverwijzingen bij kinderen met een 'risicoscore' in de interventieconditie en het percentage doorverwijzingen bij kinderen met een 'risicoscore' in de controleconditie verschilden niet significant van elkaar.

Tenslotte worden in hoofdstuk 8 de resultaten geïntegreerd en de belangrijkste uitkomsten van het onderzoek samengevat en geïnterpreteerd. Hierbij worden de methodologische beperkingen van de studies beschreven en worden aanbevelingen voor onderzoek, praktijk en beleid gedaan.

De belangrijkste conclusies luiden dat de BITSEA goede psychometrische eigenschappen heeft. De Probleemschaal heeft betere psychometrische eigenschappen dan de Competentieschaal, hoewel beide schalen goed onderscheid maken tussen kinderen met en kinderen zonder problemen. De KIPPPI, die in de controleconditie werd gebruikt ('zorg zoals gebruikelijk'), heeft ook goede psychometrische eigenschappen, maar afkappunten om kinderen met problemen te signaleren werden tot nu toe niet gebruikt in de huidige onderzoekssetting. De BITSEA heeft verscheidene sterke punten: 1) de BITSEA is een relatief korte vragenlijst, waarbij het invullen en het verwerken maar weinig tijd in beslag neemt; 2) De BITSEA kan gescoord worden en er zijn afkappunten beschikbaar, die aangeven of er mogelijk sprake is van probleemgedrag en/of een achterstand bij competenties en die toegepast kunnen worden in de praktijk (deze werden vastgesteld in dit proefschrift); en 3) de BITSEA is

een internationaal toegepaste vragenlijst en daardoor kan de prevalentie van psychosociale problemen tussen verschillende landen vergeleken worden.

Het lijkt erop dat kinderen een jaar na het consult minder psychosociale problemen ondervonden wanneer de BITSEA was ingevuld, dan wanneer 'zorg zoals gebruikelijk' werd toegepast. We raden aan deze resultaten in toekomstige studies te repliceren. Deze studie ondersteunt de BITSEA als een bruikbaar instrument voor de Nederlandse preventieve jeugdgezondheidszorg waarmee, in een vroeg stadium en op jonge leeftijd, psychosociale problemen kunnen worden gesignaleerd.

Dankwoord

En dan nu het gedeelte van mijn proefschrift waar ik al een hele tijd naar uitkijk om te schrijven; het dankwoord! Gedurende mijn gehele promotietraject ben ik mensen erg dankbaar geweest voor hun bijdrage aan mijn proefschrift, dat heb ik ze hopelijk voldoende laten weten, maar nu dan einderlijk 'zwart op wit'.

Allereerst, mijn promotor **prof.dr. Hein Raat**. Hein, ik heb veel van je geleerd. Je vond me niet (altijd) een natuurtalent, maar ik denk dat er uiteindelijk toch een heel mooi proefschrift ligt. Bedankt voor de stickertjes, in een poging om me door de laatste loodjes heen te helpen, het heeft gewerkt!

Daarnaast uiteraard **dr.ir. Wilma Jansen**. Wilma, als senior onderzoeker bij de GGD was jij als dagelijks begeleider en co-promotor betrokken bij mijn onderzoek. Bedankt voor je hulp bij het toepassen van wetenschappelijk onderzoek in de praktijk, het geven van feedback op mijn stukken, je vertrouwen en je steun. Vaak waren de overleggen ook gewoon heel gezellig. Ik heb veel van je geleerd en daar ben ik je heel dankbaar voor. Leuk dat ik in mijn nieuwe functie nog steeds met je te maken heb.

De leden van de kleine commissie, **prof.dr. Verhulst**, **prof.dr. Middelkoop** en **prof.dr. Arends**, en de leden van de **grote commissie** wil ik bedanken voor hun interesse, tijd en de aandacht die ze aan mijn proefschrift besteed hebben.

Dit proefschrift had nooit tot stand kunnen komen zonder medewerking van de **jeugdgezondheidszorgorganisaties**, CJG Rijnmond, de Zellingen, (toen nog) Careyn en Opmaat. Graag wil ik de **professionals** van deze organisaties hartelijk danken voor hun inzet en hun bereidheid te werken met een nieuw instrument. Het is erg leuk om nu te ervaren dat de BITSEA geïmplementeerd gaat worden bij CJG Rijnmond en de Zellingen!

Dear **Alice Carter**, first of all, thank you so much for granting us permission to use the BITSEA in our study. But besides that, I am thankful that you were willing to be a co-author on my papers, I learned a lot from your feedback. I had the honour and pleasure to meet with you several times in the Netherlands. You are doing very interesting work and you always talk about it with such passion; your enthusiasm is contagious. We had real good times with you here.

Beste co-auteurs: **Carolien de Haan**, **Jan van der Ende**, **Cathelijne Mieloo**, **Janne Visser**, **Tamara van Batenburg**, **Nicolien van Sprang**, heel erg bedankt voor jullie bijdrage aan de publicaties. Jullie expertise was onmisbaar.

Ik wil al mijn collega's van MGZ heel erg bedanken voor de leuke tijd die ik er gehad heb. De **statistici** (lange tijd mijn buren) waar ik vaak even kon binnenlopen met een vraag, de **databeheerders** (voor het meehelpen opzetten van mijn website en online vragenlijst) en de dames van het **secretariaat**, heel erg bedankt voor jullie hulp. Mijn collega's (van Jeugd) waren een grote steun tijdens mijn promotietraject en stonden gelukkig ook garant voor een hoop gezelligheid; er was altijd wel íemand die naar het DE café wilde: **Mirjam S**, **Mirjam vB**, **Amy**, **Vivian**, **Esther**, **Rienke**, **Teun**, **Selma**, **Anne W**, **Ilse**, **Angela**, **Suzanne**, **Anne R**, **Karen**, en **Ineke**, bedankt! Helaas heb ik na een tijd afscheid moeten nemen van **Linda** als kamergenoot, het was altijd erg gezellig. Voor haar in de plaats kwamen **Miriam** en **Daan**, en later ook **Hilde** en **Steffie**... ik weet niet eens waar ik moet beginnen met vertellen hoe leuk we het hebben gehad, misschien is het ook wijzer om dat gewoon níet te doen. Als ik zeg 'chokotoff', dan omvat dat bijna alles. Ik wil jullie in ieder geval heel erg bedanken voor jullie humor en steun. Mijn overbuurvrouw op de gang toen we nog in het Ae-gebouw zaten, en die er ook altijd voor dag en dauw was; **Amy**. Ons bijkletsmomentje om 7 uur 's ochtends was altijd een erg goed begin van de dag, ook jij bedankt, heel fijn dat ik sommige dingen bij je mocht afkijken. **Vivian**, jij bracht een hoop extra gezelligheid op de afdeling. Binnen de kortste keren waren we vriendinnen en reden we met een volgeladen auto naar Bordeaux voor een weekje vakantie in de tent. Naast alle lol die we gehad hebben, heb je me ook bepaalde inzichten gegeven waar ik veel aan gehad heb. Bedankt voor je vriendschap.

Ik was aangenomen bij MGZ en daarna bleek dat ik ook een werkplek bij de GGD kreeg. Die had ik niet zo zien aankomen en ik zat er op dat moment ook niet persé heel erg op te wachten. Uiteraard ook de GGD collega's: **Netty**, **Floor**, **Sarah**, **Vincenzo**, **Alice**, **Cathelijne**, **Carolien**, **Nandani**, **Nathalie**, **Dick**, **Bram**, **Jelle**, **Marieke**, **Irene**, **Petra**, **Anke**, **Annelies**, **Shane**, **Mounir**, **Joke**, **Anja**, **Frouwkje**, **Ralda**, **Judith**, **Tamara** en **Marisa**, heel hartelijk bedankt voor de leuke tijd die ik op de GGD gehad heb, want wat wás ik achteraf blij met die werkplek. De afwisseling van een wetenschappelijke omgeving op MGZ en een meer praktijkgerichte omgeving op de GGD was erg leuk en leerzaam. **Irene**, mijn steun en toeverlaat op logistiek gebied bij de opzet van het onderzoek, was daar bij onze eerste ontmoeting heel ontnuchterend over: ik probeerde er nog iets van te maken door te zeggen "luxe hè, 2 werkplekken", waarop zij zei "het is maar wat je luxe noemt". De toon was gezet en het resulteerde in een fijne samenwerking, een hoop gezelligheid en een mooie vriendschap. Op logisitiek gebied wil

ik ook **Marc** en **Diana** heel hartelijk danken. Ze bleven maar lege enveloppen naar mijn kamer brengen en zaten uiteindelijk met diezelfde enveloppen weer in hun maag, nu gevuld met vragenlijsten, als ze verstuurd moesten worden èn ingevuld weer terug kwamen. Ook wil ik graag **Engelien** bedanken voor het vinden van 'onvindbare' publicaties, en altijd súper snel! **Carolien**, we hebben gekke namen geëvalueerd, we hebben gelachen om 'nailed it', we hebben sushi gegeten, we hebben vooral heel veel geknutseld, o ja, en we hebben ook als onderzoekers samengewerkt aan de peuter JMR. Een lange tijd heb ik met jullie, **Carolien**

en **Irene**, op de kamer gezeten, waardoor overwerken echt niet erg was (en eerlijk gezegd ook gewoon noodzakelijk, juíst wegens alle gezelligheid). Bedankt voor jullie hulp (praat me niet meer van sjablonen of anderszins), zonder jullie was het niet gelukt! Op de GGD heb ik in verschillende kamers gezeten, in verschillende samenstellingen; het begon naast de SOA-poli op de kamer met **Sarah**, **Netty**, **Floor**, en **Vincenzo** (en later ook **Alice**) – met **Cathelijne**, **Nandani** en **Carolien** in de aansluitende kamer. Lieve collega's bedankt voor dit mooie begin van mijn promotietraject. Het was grote lol! Uiteindelijk heeft het geleid tot een heus pubquiz-team (FAN-VIS, compleet met logo, t-shirts en website), waarvan **Floor** de kapitein was en ik de coach – ik kan echter niet zeggen dat de vele oefenrondes nu echt geleid hebben tot een succes in de pub (hoewel, het bier en de wijn waren lekker). **Cathelijne**, fijn dat we gedurende onze onderzoeken zo met elkaar konden optrekken; de studies hadden veel raakvlakken (de methodes en statistieken waren namelijk net zo achterhaald ☺), en daardoor begrepen we elkaar als geen ander. We konden zelfs samen naar Stockholm voor een congres… "èn ballet!". Bedankt dat ik zoveel bij je mocht afkijken. Het was fijn dat we elkaar op het laatst om beurten gerust konden stellen.

Vrienden en **(schoon)familie** heel erg bedankt voor jullie interesse in mijn onderzoek. Jullie hebben soms (heus niet zo vaak) wat gemopper van mij moeten aanhoren, maar ook de successen kon ik gelukkig met jullie delen.

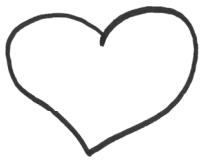
Ditta en **Denise**, ik vind het zó fijn dat jullie naast me staan tijdens de verdediging. Ik hecht enorm veel waarde aan jullie vriendschappen. Ditta, onze vriendschap ontstond op de HRBP-kamer. Onlangs gaf jij me een laatste duwtje een heel andere richting op. Jij hebt je excuses ervoor aangeboden, maar ik ben je heel erg dankbaar! Denise, onze vriendschap ontstond onder erbarmelijke omstandigheden. Heel bijzonder dat daar dan toch iets moois uit voort kan komen. Ik moet toegeven; niet alles was verkeerd, zie ook mijn 9e stelling – we kunnen er maar beter om blijven lachen. Lieve vriendinnetjes, bedankt dat jullie er altijd voor me zijn.

Tom, uiteindelijk is het iets anders gelopen, dan dat we van te voren gedacht hadden. Je bent een heel grote steun voor me geweest tijdens mijn onderzoek – zowel praktisch als emotioneel. Heel erg bedankt dat je zoveel voor me hebt gedaan, en nog steeds voor me doet. Ik ben heel dankbaar dat je nog steeds in mijn leven bent en ik hoop dat dat zo blijft.

Lieve **ouders** en **zusje**, wat ben ik blij dat ik altijd op jullie steun kon rekenen, onvoorwaardelijk en in elke vorm. **Mam**, ik heb je vanaf het begin altijd heel officieel op de hoogte gehouden van mijn vorderingen, jij was(/bent) immers mijn grootste fan. Bedankt voor je aanbod om de omslag van mijn proefschrift vorm te geven, maar uuh, we houden het gewoon maar bij het vakantie-boekje ☺.

Lieve **opa** en **oma**, ik ben heel blij dat jullie bij mijn verdediging kunnen zijn. Ik zal de knijpers op de sokken nooit vergeten en het is écht niet zo moeilijk om bij de speeltuin te komen hoor, gewoon 'hup hup, hup hup hup'. Bedankt voor jullie betrokkenheid bij mijn leven, in het begin wat intensiever dan nu, maar altijd met liefde.

Baldwin, het laatste stuk van mijn promotietraject heb ik met jou gedeeld. In het begin nog onzeker of je wel een plekje in mijn dankwoord zou 'verdienen', maar intussen zijn we wat fases verder en zijn **jij** en **Jasper** en **Tessa** niet meer weg te denken uit zowel mijn dankwoord als mijn leven. Ik hou van jullie.



About the author

CURRICULUM VITAE

Ingrid Kruizinga was born 7 November 1985 in Voorburg, the Netherlands. She completed secondary school at "Erasmus College" in Zoetermeer, after which she started a bachelor Psychology at Leiden University in 2004, and completed this in 2007. Ingrid was admitted to the two-year Research Master program of Leiden University, and she specialized in Clinical and Health Psychology. She obtained her Research Master's degree in 2009. In August 2009, she started as a junior researcher at the Department of Public Health of the Erasmus Medical Center in Rotterdam. Until 2013 she worked on evaluating instruments for the early detection of psychosocial problems in toddlers, as described in this thesis.

Currently, Ingrid is working as a researcher for the municipality of Rotterdam on several research projects, among which the decentralization of youth health care from national to local government.

PUBLICATIONS

2011

Kruizinga I, Jansen W, Carter AS, Raat H.(2011) Evaluation of an early detection tool for social-emotional and behavioral problems in toddlers: The Brief Infant Toddler Social and Emotional Assessment - A cluster randomized trial. BMC Public Health 11: 494. doi:10.1186/1471-2458-11-494

2012

Kruizinga I, Jansen W, de Haan CL, van der Ende J, Carter AS, et al. (2012) Reliability and Validity of the Dutch Version of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). PLoS ONE 7(6): e38762. doi:10.1371/journal.pone.0038762

Kruizinga I, Jansen W, de Haan CL, Raat H (2012) Reliability and Validity of the KIPPPI: An Early Detection Tool for Psychosocial Problems in Toddlers. PLoS ONE 7(11): e49633. doi:10.1371/journal.pone.0049633

2013

Kruizinga I, Jansen W, Mieloo CL, Carter AS, Raat H (2013) Screening Accuracy and Clinical Application of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). PLoS ONE 8(8): e72602. doi:10.1371/journal.pone.0072602

2014

Kruizinga I, Visser JC, van Batenburg-Eddes T, Carter AS, Jansen W, et al. (2014) Screening for Autism Spectrum Disorders with the Brief Infant-Toddler Social and Emotional Assessment. PLoS ONE 9(5): e97630. doi:10.1371/journal.pone.0097630

Brosschot JF, Geurts SAE, Kruizinga I, Radstraak M, Verkuil B, Quirin M, Kompier MAJ (2014) Does unconscious stress play a role in prolonged cardiovascular stress recovery? Stress and Health 3: 179-187.

Submitted

Kruizinga I, Jansen W, Van Sprang NC, Carter AS, Raat H. The effectiveness of the BITSEA as a tool to early detect psychosocial problems in toddlers, a cluster randomized trial. Submitted.

PHD PORTFOLIO

Name PhD student:	Ingrid Kruizinga
Erasmus MC Department:	Public Health
PhD period:	2009-2013
Promotoren:	Prof.dr. H. Raat
Copromotor:	Dr.ir. W. Jansen

		Year	Workload
	Courses/workshops		
-	Causal Inference	2010	0.7 ECTS
-	Methods Health Services Research	2010	0.7 ECTS
-	Primary & Secondary Prevention Research	2010	0.7 ECTS
-	Social Epidemiology	2010	0.7 ECTS
-	Methodologie van Patiëntgebonden Onderzoek en Voorbereiding Subsidieaanvragen	2010	8 hours
-	How to write a grant proposal (ZonMw)	2013	8 hours
	Presentations/posters		
-	Jeugd in Onderzoek, Nieuwegein, the Netherlands "VEST: Validatie en Evaluatie van een ScreeningsinstrumenT voor peuters" (poster presentation)	2010	12 hours
-	Jeugd in Onderzoek, Nieuwegein, the Netherlands "Evaluatie van het gebruik van een nieuw screeninsinstrument in de praktijk - BITSEA" (poster presentation)	2011	12 hours
-	NCVGZ, Amsterdam, the Netherlands "Betrouwbaarheid en validiteit van de Brief Infant-Toddler Social and Emotional Assessment (BITSEA)" (oral presentation)	2012	12 hours
-	Research seminar Department of Public Health, Erasmus MC, Rotterdam, the Netherlands "The evaluation of two early detection instruments for the measurement of psychosocial problems in toddlers" (oral presentation)	2012	5 hours
-	Jaarcongres Jeugdgezondheidszorg, Ede, the Netherlands "De nauwkeurigheid van de BITSEA en SDQ in het signaleren van psychosociale problemen" (poster presentation)	2012	12 hours
-	Cephir seminar Prevention, Erasmus MC, Rotterdam, the Netherlands "De preventie van psychosociale problemen door vroege signalering in de jeugdgezondheidszorg" (oral presentation)	2013	5 hours
-	European Congress of Psychology, Stockholm, Sweden "BITSEA and SDQ: Screening for psychosocial problems in preventive child health care" (oral presentation)	2013	48 hours
	Seminars/symposia		
-	Seminars at the Department of Public Health, Erasmus MC, Rotterdam	2009-2013	2 ECTS
-	CEPHIR seminars, Rotterdam,	2009-2013	0.5 ECTS
-	Seminar GGD Nederland, Utrecht 'Non-respons'	2009	4 hours
-	Jaarcongres Jeugdgezondheidszorg, Ede, the Netherlands	2009	8 hours

		Year	Workload
-	NCVGZ, Rotterdam, the Netherlands	2010	8 hours
-	Autisme bij jonge kinderen, Utrecht, the Neterlands	2010	8 hours
-	Jaarcongres jeugdgezondheidszorg, Ede, the Netherlands	2010	8 hours
-	DWARS conferentie Social Marketing	2011	4 hours
-	Social Media seminar, Rotterdam, the Netherlands	2012	4 hours
	Teaching		
-	Supervision of Erasmus MC medical students in writing assignment of 'Effectiveness of the NIDCAP intervention in newborns'	2009	8 hours
-	Supervision of Erasmus MC medical students in writing assignment of 'Effectiveness of early detection of psychosocial problems in young children (0-4 years old)	2010	16 hours
-	BITSEA training youth health professionals	2010	16 hours
-	Supervision of master student 'Developmental Psychology' Leiden University. Thesis title: "Early detection of psychosocial problems in Dutch toddlers. The psychometric properties of the Dutch version of the BITSEA"	2011-2012	45 hours
-	Supervision of master student 'Developmental Psychology' Leiden University. Thesis title: "Differences in perceived user-friendliness between the BITSEA and the KIPPPI"	2012-2013	45 hours