The Identification by Dutch Preventive Child Health Care of Children with Psychosocial Problems:
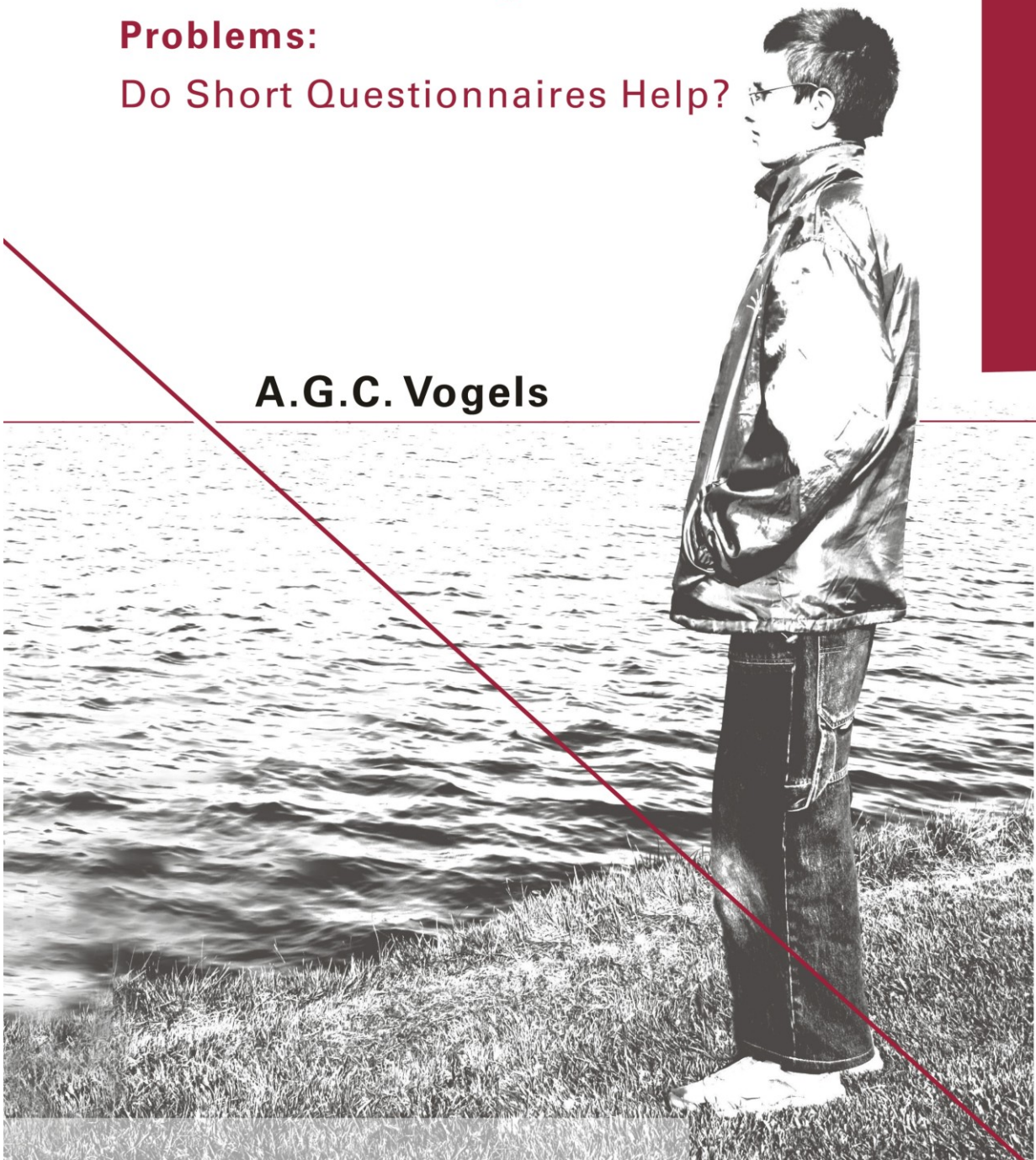
Do Short Questionnaires Help?

A.G.C. Vogels

The Identification by Dutch Preventive Child Health Care of Children with Psychosocial Problems:
Do Short Questionnaires Help?

A.G.C. Vogels

RIJKSUNIVERSITEIT GRONINGEN

The Identification by Dutch Preventive Child Health Care of Children with Psychosocial Problems:
Do Short Questionnaires Help?

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
woensdag 23 januari 2008
om 16.15 uur

door

Antonius Gerardus Cornelis Vogels
geboren op 25 november 1950
te Gemert

Preface

Writing this thesis was only possible with the help and support of many people and I would like to express my gratitude to all of them.

First of all, of course, Menno Reijneveld and Pauline Verloove. Pauline, thanks for pushing me, for a long time, but finally with success. Thanks, also, for your continuing support after my decision not to write a thesis on your favorite subject. Menno, my appreciation for your role in writing this thesis is as immeasurable as the number of critical and supportive comments you made on drafts I sent to you. I hope very much that we will be able to continue our collaboration.

Also, I would like to thank Prof. dr. R. Sanderman, Prof. dr. P.J.J. Sauer and Prof. dr. F.C. Verhulst for their willingness to read this manuscript and to allow it to be defended. as a thesis. Prof. Dr. F.C. Verhulst was involved in some of the studies reported in this thesis. Thanks for that, too.

The manuscripts which are part of this thesis were not written by me alone. To all authors and co-authors of these manuscripts: thanks! I enjoyed working with you. The same goes for all those many, many colleagues at TNO Quality of Life, who are not mentioned as (co-)author, but without whose efforts the studies reported in this thesis would not have been possible.

Also, many people outside TNO Quality of Life must be mentioned. Thanks to all those youth doctors, youth nurses and secretarial staff in Dutch Preventive Child Health Care who participated in our studies. Your willingness to co-operate in each of the studies was essential.

Most of the studies reported were done in close co-operation with the Landelijke Werkgroep Signalering van Psychosociale Problematiek bij Jongeren. I want to thank all members of this working group for the cordial collaboration. I am convinced that your role in the development of an evidence based detection of children with psychosocial problems cannot be overestimated. Without you, this thesis would not have been written.

Ton, thanks for what you are to me and for what you did for this manuscript, with – in your own words – 'unlimited patience'.

Ton Vogels

Content

1 Introduction and outline of this thesis

A.G.C. Vogels

1.1 Introduction

This thesis focuses on the early detection by the Dutch Preventives Child Healthcare (PCH) of children with psychosocial problems and examines in particular whether short questionnaires can improve early detection.

Most Dutch children are doing well. A recent study compared the well-being of children in 21 rich countries and concluded that children in the Netherlands were better off than those in other countries.[1] The Health Behavior in School Children study also concluded that, on average, Dutch children did well and were more satisfied with their lives than children in other European countries.[2] However, such global findings do not imply that all children are doing fine and that all children are doing well and are feeling happy. On the contrary, many studies have shown that a great number of children suffer from serious psychosocial problems. Brugman et al.[3] studied the prevalence of emotional and behavioral problems in children aged five to fifteen years, using the Child Behavior Checklist (CBCL) and the Youth Self Report (YSR) and found prevalences ranging between 6 and 9%. A more recent study by Zeil et al.[4] studied the age group from 14 months to twelve years of age, using the CBCL and the Infant and Toddler Emotional Assessment Scale (ITSEA) and found prevalences ranging between 2% (very young girls) and 8% (boys aged five or six years).

These figures are based on studies using a cut-off point, which are, inevitably, to a certain extent arbitrary. However, there are also other indicators showing that many Dutch children and adolescents suffer from serious psychosocial problems, hampering normal, healthy development[5]: in 2004 there were about 33,000 new admissions to the youth mental-healthcare services; 52.000 new voluntary admissions to Bureau Jeugdzorg (Youth Welfare Work, BJZ) and 7,000 new admissions to youth rehabilitation centers. The total number of children under the age of 19 receiving treatment in youth mental healthcare services was 84,000. An additional 57,000 youngsters were receiving treatment at centers for justitional youth care. Furthermore, in that same year, more than 33,000 handicapped youngsters under the age of 25 were receiving a disability allowance because they were not fit to work (Wajong). Of all cases newly enrolling for this allowance, 39% were doing so on the basis of some psychological – not intelligence related – handicap.

In short, psychosocial problems can seriously hamper children's healthy development and are often a burden for parents and other family members. In many cases, these problems tend to be persistent and they do not disappear automatically,[6] possibly because they may have been caused by physical and genetic factors or by experiences in very early childhood. They may have long-term negative consequences, like academic under-achievements, substance abuse, the need for referral to mental healthcare or professional counseling, problems at school and clashes with the law.[7-12]

In the Netherlands PCH has, by law, the duty to identify children with psychosocial problems and, if necessary, to assure that they are given adequate care. This is an important aspect of its work as research has shown that many children with serious problems do not receive any form of help at all.[13] This despite the fact that research has also shown that early detection of children with emotional and behavioral problems, if followed by adequate treatment, significantly improves their prognosis.[14,15]

1.2 PCH in the Netherlands: a short description

Reijneveld et al.[16] describe Dutch PCH as that part of the healthcare system that focuses on the promotion, protection and safeguarding of health, growth and physical and mental development. Its target group are all children, between zero and nineteen years of age. It distinguishes itself from other parts of the healthcare system in that it offers its services pro-actively, not waiting until the moment that a problem may arise. Traditionally PCH is offered by two different types of organizations: well-baby clinics, (Consultatiebureaus, CB), mostly a service offered by the 'Home-Help Organizations' (Thuiszorgorganisaties), targeting (parents of) children aged 0 to 4, and PCH from 4 to 19, a service offered by regional public health centers (Gewestelijke Gezondheidsdiensten, GGD). PCH offers both individual and collective services, in a variety of disciplines: specialized youth doctors, community and youth nurses, epidemiologists, dieticians, speech therapists, epidemiologists and sometimes psychologists. Children are invited for a standard health examination at specific intervals, 15 times between 0 and 4 years of age and 3 times when they are at primary and secondary schools.

The main services are defined in PCH's Basic Working Package.[17] Some of the services are a standard, uniform part of this working package and have to be provided systematically to all children involved, in a standardized way. The main services are:

1. Monitoring and identification of problems
2. Estimating the need for (extra) care
3. Screening and immunization
4. Health promotion, advice, instruction and support, aiming to improve health related behavior
5. Handling of health risk
6. Supporting the (mental) healthcare system

The identification of children with psychosocial problems is part of the service mentioned under 1 and belongs to the uniform part of PCH's Basic Working Package. When a child is identified as having problems, PCH should, of course, carefully estimate the type of care needed in this child's case. Identifying children with problems is, primarily, a service offered to individual parents and children. Additionally, it should contribute to the collectively offered monitoring service, in which PCH systematically tries to monitor the health of the population under its care and advises local and regional health authorities on the development of adequate healthcare policies.

1.3 The Term 'Psychosocial Problems'

What does the term "psychosocial problems" mean? It is a common term in Dutch PCH, but its definition presents difficulties. In 1985 Vogels et al.[18] contacted representatives of all kinds of services working with young people and asked them to indicate the most important psychosocial problems facing children and adolescents. The answers varied enormously: problems with parents or peers, homosexuality, fear of failure, unemployment, having no future perspective, divorce of the parents and so on.

Pubmed does not use the phrase as one of its standard keywords. The Pediatric Handbook Online (http://www.rch.org.au/paed_handbook/index.cfm?doc_id=1571, date May 16, 2007) has an entry titled "psychosocial problems" but does not offer a clear definition. Instead it lists a range of problems, varying from anxiety disorders, family relationships difficulties, psychosis, ADHD, sleep problems, learning difficulties, child abuse and neglect and – remarkably – death of a child.

In 1999 the study group 'Monitoring Stations in Dutch PCH' organized a meeting of experts to discuss prevention of psychosocial problems among children and adolescents by PCH. Leading experts in the field of PCH and child and adolescent psychiatry contributed to this meeting. One of the issues discussed was how to define the term psychosocial problems in this specific context. Those present agreed that a clear distinction should be made between factors contributing to or causing the development of psychosocial problems and psychosocial problems themselves. Furthermore they agreed that the early detection of psychosocial problems should focus on emotional and behavioral problems and on problems at school. This definition limits the term psychosocial problems to problems in the functioning of the child itself. The Association of Dutch Regional Health Services (GGD Nederland, GGD NL) and the Dutch Institute for Care and Wellbeing (Nederlands Instituut voor Zorg en Welzijn (NIZW)) adopted a similar approach: psychosocial problems are described as problems in psychosocial functioning; they may be emotional but are often also visible in the child's behavior.[19] This definition will also be used in this thesis.

1.4 Psychosocial problems and Dutch PCH

The early detection of children with psychosocial problems has been part of the uniform part of PCH's Basic Working Package since 2002. Since its explicit inclusion, this part has attracted an increasing amount of attention, and a great deal of importance has been attached to it. This high profile was caused by several dramatic cases of child abuse and child neglect, which made it clear that Dutch institutions in mental and physical healthcare were insufficiently focused on the identification and management of children at risk. Nevertheless, the late inclusion (2002) of the early detection of psychosocial problems as a an essential part of PCH's working package, does not mean that it was totally ignored before that date.

Reijneveld et al.[20] suggest that PCH always paid some attention to psychosocial problems, but that in the beginning other issues, like clean drinking water, healthy nutrition and hygiene were more relevant and urgent, simply because the main issue then was to safeguard health and improve survival rates. They see the beginning of the 1970s as the time when interest in psychosocial problems started to increase. In 1971 the Netherlands Institute for Preventive Medicine (Nederlands Instituut voor Praeventieve Geneeskunde, NIPG), now part of TNO Quality of Life, published a report on secondary school pupils' complaints, concluding, amongst other findings, that the identification of these problems required better trained and more highly educated staff.[21] Ter Haar Romeiny-Wachter and Wassen-Van Schraven published two articles

describing a roadmap to preventive mental healthcare, in which they formulated a clear role for well-baby clinics.[22]

It is difficult to assess what exactly happened in the PCH practice at that time. Vogels et al. studied the annual reports of al PCH departments in the Netherlands, for the years 1982, 1983, 1984 and 1985.[23] This was done in preparation of a study on PCH for older adolescents. Psychosocial problems or mental health were rarely mentioned in those reports; far fewer references were found than, for example, to the number of children with head lice. Yet, at the same time the interest in psychosocial problems was clearly growing in the field. The department for PCH in South-East Drenthe asked the NIPG to carry out a large-scale study on the prevalence of psychosocial problems. This study focused on pupils in secondary vocational schools but was also to include youngsters who had already left school.[24,18] Other regional healthcare centers took similar initiatives. Using the well validated CBCL, PCH in South-East Limburg did a study aiming to establish differences in the prevalence of psychosocial problems between regions in their catchment area and between individual schools.[25,26] This was done with the explicit aim of developing a differentiated program of care. The PCH in Apeldoorn also did a study to determine the prevalence of psychosocial problems in the population falling within their catchment area.[27] By then, several centers had also introduced consultation hours where pupils and parents could seek advice, mainly focusing on psychosocial problems. Some centers decided it was worthwhile hiring people with specific expertise: psychologists or a nurse certified as a welfare worker.

All the projects mentioned above were local initiatives, focusing on different groups and or problems, using different measuring instruments and methodologies, and lacking any form of standardization. However, at a national level, interest in psychosocial problems was growing too. In 1983, the Netherlands Association for Preventive Child Healthcare (Nederlandse Vereniging voor Jeugdgezondheidszorg, NVJG) instituted a commission whose remit was to formulate guidelines for the care for adolescents. In its final report this commission stated that in this age group more attention should be paid to psychosocial problems than to problems connected with physical development.[28] This report resulted in the foundation of the Netherlands Association for Care for Adolescents (Nederlandse Vereniging voor Adolescentenzorg, NVAZ). This association started to publish a quarterly, almost exclusively directed at psychosocial development. The recommendations of the commission's report, however, were never implemented at a national level.

At governmental level interest in psychosocial problems was growing, too. In 1985 a chain of suicides in one school in the North of the Netherlands received a lot of media coverage. The regional PCH center was criticized severely for taking too few initiatives to prevent this chain of suicides. The Inspector for PCH, then part of the Health Department, took the initiative to institute a Working group on PCH and Prevention of Suicide, whose remit was to formulate guidelines for PCH which would actively contribute to the prevention of suicide in adolescents. Some of the recommendations of this commission, especially those on reinforcing the chain of care, seem very modern, but they were never actually implemented.[29]

So, for quite a while, the amount of interest shown in psychosocial problems and the action taken by local PCH departments was determined mainly by local priorities and

individual interests. At the end of the nineties, this changed, partly because the decision was taken to formulate the PCH Basic Working Package and partly, too, because PCH departments were becoming more experienced in dealing with questionnaires on psychosocial problems.

1.5 Questionnaires on psychosocial problems in Dutch Preventive Child Healthcare

By the mid-eighties nearly all departments already used short questionnaires as part of the care package offered to individual children and parents. Each PCH department used its own questionnaire which were all different, measuring different concepts, and using different items. Mostly they were about one page long, but they focused on a wide range of aspects, covering physical health, life style, and psychosocial problems. Most of these questionnaires were never validated. The questionnaires were used mainly to help parents prepare for the standard health examination: they were intended as a kind of checklist for subjects that needed attention during that examination.

Some PCH departments, however, went further and considered using questionnaires to select those children that needed to be examined in a face to face examination, especially with regard to adolescents. This led to fierce debates. Verhoeven Tjan, for example, reported on a study comparing the findings based on an unvalidated questionnaire on psychosocial problems with findings based on the consultation during a standard health examination. She found considerable discrepancies between the two methods and concluded that questionnaires, therefore, could not replace the traditional standard health examination.[30] Other studies addressed similar questions.[31,32]

Baecke reported about one of the first attempts to develop a standardized and validated questionnaire, focusing mainly on psychosocial problems.[33] He intended this questionnaire to be used as a sort of screening questionnaire, distinguishing children with problems from those without. Its validation, however, was weak: he only reported significant differences on a personality test for children between pupils with extremely high and low scores on the questionnaire's subscales. Although a number of other PCH departments did use this questionnaire, it was never adopted universally. One of the likely reasons for this is that in the beginning of the 1990s the volume of PCH services for adolescents gradually started to decrease, as financial cutbacks and changing systems of financing led many PCH centers to focus on younger age groups. Other PCH centers started to experiment with published questionnaires, e.g. the Delft Questionnaire (Delftse Vragenlijst[34]) used for some time by the PCH center in South East Limburg.

A turning point in this confusing development was the inauguration in the late nineties of the working group on the identification of psychosocial problems. It was instituted by GGD NL and officially called the National Working Group on the Identification of Psychosocial Problems in Youth (Landelijke Werkgroep Signalering van Psychosociale Problematiek bij Jongeren', LSPPJ). This working group consisted of epidemiological and sociological researchers and of school doctors working at regional PCH centers. They developed the National Checklist for Indicating Psychosocial Problems in Five-Year-Olds (Landelijke Signaleringslijst voor Psychosociale Problematiek bij Kleuters,

LSPPK). They tested its internal consistency, scale structure and validity, i.e. the extent to which it could distinguish between those receiving treatment and those not.[35] Similarly, they developed the Short Indicative Questionnaire for Psychosocial Problems among Adolescents (Korte Indicatieve Vragenlijst voor Psychosociale Problematiek bij Adolescenten, KIVPA).[36] These instruments were intended as tools to be used in the standard health examination, not as a replacement for it and not as a screening device as such. The aim was to give PCH professionals an indication of existing problems and help them focus the consultation with the parents on these problems. Having developed these questionnaires, the LSPPJ supported the implementation on a national scale: they wrote manuals, developed scoring forms and introduced methods for optical data entry so that the data collected could be used, not only in the context of individual care, but also for epidemiological purposes. Furthermore, they co-operated wholeheartedly in critical evaluations of these instruments and used the results of these studies to improve the questionnaires. Later on they developed a questionnaire for psychosocial problems affecting primary school children (vragenlijst voor PSYchosociale problematiek in de BOvenbouw van het BAsisonderwijs, PSYBOBA) and decided to participate in the evaluation of this instrument in comparison with two other instruments.

By doing so, the LSPPJ has contributed enormously towards an evidence-based system of detecting psychosocial problems in Dutch PCH and towards a reduction of the uncontrolled proliferation of unvalidated methods and instruments. Their efforts have also have led to a consensus on the criteria which must be met by questionnaires to be used by PCH.[37] First of all of course, questionnaires must be valid and reliable indicators of the phenomenon to be assessed, in this case: the likelihood of psychosocial problems being present. Also, questionnaires need to have added value: they should improve the chances of identification compared to identification based on other methods, like using information about risk indicators or carrying out an interview during the standard health examination. However, the specific PCH context demands the fulfillment of two extra criteria: PCH professionals have only a limited amount of time to spend on each individual standard health examination, which means that (paper and pencil) questionnaires have to be short, and that PCH professionals can easily work out the scores. Also, such questionnaires have to be suitable for children or parents from different backgrounds, including all socio-economic classes and all ethnic minorities. That means – again – that they cannot be very long, but also that they must be very easy to understand and that their phrasing and content must be acceptable for people from different backgrounds. Finally and preferably, the questionnaire must help prepare the respondents (parent and/or child) for the standard health examination. This being the case, the questionnaire can be used to structure the conversation between parent/child and PCH professional.

1.6 Early detection of psychosocial problems by Preventive Child Healthcare

What PCH has to do, essentially, is to determine which children are in need of special attention as a result of psychosocial problems. This sounds very much like screening

and many of the methods used in this thesis are also often used in the evaluation of screening procedures. The criteria for questionnaires that must be met, for example, correspond to those discussed by Wilson and Junger.[38] Yet, a closer inspection of the concept of screening shows that early detection of psychosocial problems cannot be viewed as a proper form of screening. Wilson and Jungner[38] define screening as 'the presumptive identification of unrecognized disease or defect by the application of tests, examinations or other procedures which can be applied rapidly". They discussed three criteria for the evaluation of screening tests (validity, reliability and 'yield' (i.e. the number of newly detected cases)) and formulated ten principles that must be born in mind when planning screenings:

1. The condition sought should be an important health problem
2. There should be an accepted treatment for patients with recognized disease
3. Facilities for diagnosis and treatment should be available
4. There should be a recognizable latent or early symptomatic stage
5. There should be a suitable test or examination
6. The test should be acceptable to the population
7. The natural history of the condition, including development from latent to declared disease should be adequately understood
8. There should be an agreed policy on whom to treat as patients
9. The cost of case-finding should be economically balanced in relation to possible expenditure on medical care as a whole
10. Case-finding should be a continuing process

In time, these principles have grown to be accepted as criteria that must be met for a screening program to be acceptable. Recently, they were adapted and updated.[39-41] Clearly, this definition is not suitable for the identification of children with psychosocial problems. The main reason for this is that psychosocial problems, although they may be related to psychiatric disorders, cannot be seen as defined diseases or defects. Furthermore, although there is evidence that high-quality early detection of psychosocial problems improves the prognosis, there is not always an accepted treatment, there is no clear latent stage and much remains unknown about the natural history of – the diversity of – psychosocial problems.

Nevertheless, the identification of children with psychosocial problems is part of the Dutch PCH's Basic Working Package, because PCH's overall aim is to monitor and safeguard the healthy development of all children in the Netherlands. If this aim is to be realized, the identification of psychosocial problems is essential, because psychosocial problems may be indicators of the existence of psychiatric disorders needing professional care and because psychosocial problems by themselves can be a risk factor for a healthy development, for example when parents are lacking in the skills needed to cope with those problems, in which case the interaction between parents and child may deteriorate. Another possibility is that such problems may be the underlying cause of physical problems, such as headaches, repeated infections, abdominal pains et cetera.

The early identification by PCH of children with psychosocial problems is a first step, that identifies those children who need more attention because of the likely existence of psychosocial problems. In other words, it is a procedure which switches on an alarm, thus calling for attention. After this signal, PCH professionals have to assess by means

of other procedures a) whether any psychosocial problem is indeed present, b) what the nature of the problem is, c) how serious the problem is, and d) whether parent and/or child are in need of support. Therefore, early detection needs to be supplemented with some form of assessment that enables PCH professionals to determine whether child and/or parent need extra care, and if so, what kind of care.

Several studies have shown that PCH identifies many children as having psychosocial problems. Reijneveld et al. for example found that PCH, using only information on file and the consultation during the standard health examination, identifies about 9% of children aged 14 months to 4 years of age as having problems.[42] In older age groups this percentage is considerably higher. Brugman et al. reported that among children aged 5 to 15, 25% had been identified as having at least one problem.[42] Zeijl et al. reported percentages varying from 11% among 11-months old children to 28% among children aged five to six years[4] The same studies, however, also strongly suggest that the identification of problems by PCH is far from perfect. When comparing the identification by PCH with the CBCL-score of the children involved, Brugman et al.[43] reported that the identification of problems was 6 times more likely in children with a clinical CBCL Total Problem Score. However, PCH identified no psychosocial problems in 43% of the children with a clinical CBCL. Also, PCH suspected that such problems existed among a large number of children, even though CBCL data suggested few or no problems. Reijneveld et al.[44] and Zeijl et al.[4] have reported similar discrepancies.

This in itself does not mean that PCH problem identification is wrong. CBCL scores are based on behaviors, moods and problems perceived and reported by parents. PCH may see different aspects of the child's functioning and research has shown that different informers' perceptions should be considered as supplementary and worthwhile information, both in their own right and in combination.[45,46] Yet, both the concurrent and the predictive validity of the CBCL and the YSR as instruments to detect emotional and behavioral problems has been proven over and over again. This means that a clinical CBCL or YSR score should be taken seriously and that the discrepancies between PCH findings and CBCL indications are strong signals that identification by PCH may be in need of improvement.

There are several ways in which this improvement might be achieved. Wiefferink et al.[47] studied the effectiveness of using a detailed protocol, in combination with a specific training. Clearly, time may be a critical factor for a high-quality identification of psychosocial problems to be achieved. Another method, tried in the past but never explicitly evaluated, is adding specific expertise to that already present in PCH, for example introducing clinical psychologists and/or social workers. In PCH for younger children, a number of rather intensive methods are being introduced, involving protocol-based home visitation (Samen Starten, Project Oké). Several studies have given a strong indication that these approaches are effective, improving both identification and offering easy opportunities for interventions.[48] Durlak & Walls[14,15] studied the effectiveness of interventions in the field of mental health. One critical factor they identified for the effectivity of interventions was the use of validated questionnaires to identify the children to which those interventions should be directed. This method is clearly in line with standard methods used by Dutch PCH in the past. The extent to which

questionnaires can improve the identification of children with psychosocial problems is the main focus of this thesis.

1.7 Research questions and outline of this thesis

The following questions will be answered in this thesis:

1. Can differences in the number of children identified as having psychosocial problems by individual health professionals be explained by differences in the prevalence of problems or background characteristics in the groups of children they examine?
2. What are the psychometric qualities of a number of short questionnaires PCH could use to identify children with psychosocial problems, and could they improve the identification by PCH?
3. Is it possible to develop a Computerized Adaptive Test using items from questionnaires on psychosocial problems, in order to achieve a short, yet accurate assessment of the likelihood of psychosocial problems being present?

Data from several studies have shown not only discrepancies between problem identification and validated indicators of psychosocial problems, but also large differences between individual PCH professionals in the percentage of children they identify as having problems. In Chapter 2 we will assess whether these differences are larger than may be expected on the basis of random variation and whether these differences may be explained by differences in the prevalence of problems or risk factors in the subsamples examined by individual professionals.

In Chapter 3 the psychometric properties of the KIVPA, intended to be used in PCH for adolescents, will be evaluated. The scale structure and validity – in terms of sensitivity and specificity – are evaluated. Also, it will be determined whether using this questionnaire offers added value, compared to identifying children based on information about potential risk indicators.

Chapter 4 will assess the psychometric properties of one of the questionnaires developed by the LSPPJ, the LSPPK. The LSPPK is intended to be used in PCH for 5 and 6 years old. The methods used in this chapter are comparable to those used in chapter 3.

The PSC, the Pediatric Symptom Checklist, is widely used in the USA for the detection of children with problems in all kinds of settings. Several studies have shown strong psychometric properties. Chapter 5 describes a study evaluating the psychometric properties and its added value for Dutch PCH for children in the second phase of primary education. In this chapter the concept of added value is extended, compared to the definition used in the previous two chapters. We will not only determine whether the PSC allows for a better identification than potential risk indicators, but also whether it improves the identification based on the clinical judgement of the PCH professional, after having examined the child.

In Chapter 6 a randomized comparison of three questionnaires is described, the SDQ, the PSC and the PSYBOBA, developed by the LSPPJ. The aim of this study is to determine which of these questionnaires would be most suitable for the identification of

children with problems in the 7 to 12 age group. The questionnaires will be compared in terms of scale structure, validity and the extended added value they may offer to PCH. Also, their suitability, both for parents and PCH professionals, will be compared.

The Strengths and Difficulties Questionnaire (SDQ) was developed in the United Kingdom.[49,50] The quality of the instrument has been proven in different countries and cultures. There is evidence that the quality of the Dutch version is high, too.[51,52] Chapter 7 presents a detailed evaluation of four different SDQ-based classification methods, comparing their validity and added value for the identification of children with problems in the second phase of primary education by PCH professionals.

Questionnaires to be used by PCH must, of course, offer valid and reliable indicators for the existence of psychosocial problems. Due to the strong time limitations within which PCH has to work, they must meet another criterion: they have to be short and easy to score. Short questionnaires are, however, inherently less reliable than longer questionnaires.[53] A lack of reliability means, by definition, a greater chance of mis-identification. Item Response Theory (IRT) provides a way out of this dilemma, offering a method with which assessment using only a few items can result in a highly accurate measurement. Chapter 8 will describe a study assessing whether the items from four questionnaires may be used for an IRT-based computerized adaptive test.

In Chapter 9 the main findings of this thesis are summarized and discussed.

References

1. UNICEF. Child poverty in perspective: an overview of child well-being in rich countries. Innocenti Report Card 7. Florence: Unicef Innocenti Research Center; 2007.

2. Curry C, Roberts C, Morgan A et al. Young people's health in context. Health Behavior in School-aged Children (HBSC) study: international report from the 2001/2002 survey. Geneva: WHO Regional Office for Europe; 2004.

3. Brugman E, Reijneveld SA, Hollander-Gijsman ME, Burgmeijer RJF, Radder JJ. Psychosociale gezondheid van de jeugd tussen vijf en vijftien jaar. *GGD-Nieuws* 2000;(February 2000):19-23.

4. Zeijl E, Crone MR, Wiefferink CH, Keizenkamp S., Reijneveld SA. *Children in the Netherlands [in Dutch]*. The Hague: SCP; 2005.

5. Besseling J, Hagen B, Andriessen S et al. Toename gebruik ondersteuning voor jongeren met een gezondheidsbeperking, Den Haag: Ministerie van Sociale Zaken en Werkgelegenheid; 2007.

6. Verhulst FC, Van der EJ. Six-year stability of parent-reported problem behavior in an epidemiological sample. *J Abnorm Child Psychol* 1992 December;20(6):595-610.

7. Ferdinand RF, Blum M, Verhulst FC. Psychopathology in adolescence predicts substance use in young adulthood. *Addiction* 2001 June;96(6):861-70.

8. Visser JH, Van der EJ, Koot HM, Verhulst FC. Continuity of psychopathology in youths referred to mental health services. *J Am Acad Child Adolesc Psychiatry* 1999 December;38(12):1560-8.

9   Verhulst FC, Koot HM, Van der EJ. Differential predictive value of parents' and teachers' reports of children's problem behaviors: a longitudinal study. *J Abnorm Child Psychol* 1994 October;22(5):531-46.

10   Heijmens Visser JH, Van der EJ, Koot HM, Verhulst FC. Predictors of psychopathology in young adults referred to mental health services in childhood or adolescence. *Br J Psychiatry* 2000 July;177:59-65.:59-65.

11   Ferdinand RF, Visser JH, Hoogerheide KN et al. Improving estimation of the prognosis of childhood psychopathology; combination of DSM-III-R/DISC diagnoses and CBCL scores. *J Child Psychol Psychiatry* 2004 March;45(3):599-608.

12   Koot HM, Verhulst FC. Prediction of children's referral to mental health and special education services from earlier adjustment. *J Child Psychol Psychiatry* 1992 May;33(4):717-29.

13   Verhulst FC, Ende van der J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997 July;36(7):901-9.

14   Durlak JA, Wells AM. Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *Am J Community Psychol* 1998 October;26(5):775-802.

15   Durlak JA, Wells AM. Primary prevention mental health programs for children and adolescents: a meta-analytic review. *Am J Community Psychol* 1997 April;25(2):115-52.

16   Reijneveld S.A., Verloove-Vanhorick SP, Hirasing RA. Jeugdgezondheidszorg. In: Leerdam van F, Hirasing RA, editors. *Basisboek Jeugdgezondheidszorg*. Elsevier Gezondheidszorg; 2006. p. 6-17.

17   Ministerie van VWS. Basistakenpakket Jeugdgezondheidszorg. Den Haag: Ministerie van VWS; 2002.

18   Vogels T, Andries F, Visser A. Adolescenten en Jeugdgezondheidszorg. Onderzoek onder oudere adolescenten in Zuid Oost Drenthe. Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg; 1985.

19   Blokland G. Groepsgerichte opvoedingsondersteuning. *Nederlands Tijdschrift voor Jeugdzorg* 2002;(2):97-103.

20   Reijneveld S.A. Zat het in 1901 nog niet tussen de oren? Herkenning van psychosociale problemen bij kinderen. In: Bakker W, Damen MLW, Merkx JAM, Prinsen B, editors. *Alle kinderen: Ouder- en Kindzorg tussen wetenschap en praktijk*. Assen: Van Gorcum; 2001. p. 77-82.

21   Groen M, Smit J. *Klachten van middelbare scholieren*. Nederlands Instituut voor Praeventieve Geneeskunde; 1971.

22   Ter Haar Romeiny Wachter CC, Wassen-Van Schaveren PCJ. Op weg naar preventieve geestelijke gezondheidszorg I. *TSG* 1793;51:790-3.

23   Vogels T, Visser A, Andries F. *Boven de norm - een inventarisatie van aktiviteiten vanuit de JGZ ten behoeve van oudere adolescenten*. Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg; 1987.

24   Schuurman MIM, Visser A. Adolescenten en Jeugdgezondheidszorg. *Onderzoek onder LBO-leerlingen in Zuid Oost Drenthe*. Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg; 1985.

25   Vogels T. *Psychische problemen bij leerlingen in het tweede leerjaar van het voortgezet onderwijs in Oostelijk Zuid-Limburg I. Verschillen tussen onderwijssoorten en jongens en meisjes.* Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg; 1992.

26   Vogels T. *Psychische problemen bij leerlingen in het tweede leerjaar van het voortgezet onderwijs in Oostelijk Zuid-Limburg II. Verschillen tussen gemeenten en buurten in Kerkrade.* Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg; 1992.

27   Maarsingh EJ. Signaleren van psychosociale problematiek bij adolescenten. *TSG* 1996;98(4):54-7.

28   Franken EBM, Have ten F, Kooij-Pontiers van den HA, Lodder-van der Kooij AM, Pijpers FIM. *Sociaal-geneeskundige adolescentenzorg.* 's Hertogenbosch: Nederlandse Vereniging voor Jeugdgezondheidszorg; 1984.

29   Vogels AGC. *Jeugdgezondheidszorg en suïcidaal gedrag onder adolescenten. Rapport van de werkgroep JGZ en Suïcidepreventie.* Leiden: Nederlands Instituut voor Praeventieve Gezondheidszorg - TNO; 1986.

30   Verhoeven-Tjan VBG. Kunnen vragenlijsten het periodiek geneeskundig onderzoek bij adolescenten vervangen? *TSG* 1987;(65):114-5.

31   Pijnenburg-Dewez L. Kunnen vragenlijsten het periodiek geneeskundig onderzoek bij oudste kleuters vervangen? *TSG* 1987;(65):167-8.

32   Kok C, Wezel Pv, Maarsingh EJ, Vogels T, Lako CJ. Signaleren van psychosociale problematiek bij adolescenten. Het Preventief Gezondheidsonderzoek en een vragenlijst vergeleken. *TSG* 1994;72(72):363-8.

33   Baecke JAH. Project epidemiologie in Basisgezondheidsdiensten: de ontwikkeling van een gezondheidsvragenlijst voor adolescenten. *TSG* 1987;(65):99-103.

34   Appels A. *Screenen als methode voor preventie in de geestelijke gezondheidszorg. Tevens handleiding bij de Delftse Vragenlijst.* Amsterdam: Swets & Zeitlinger; 1975.

35   Bouchier CAM, Rover de CM, Vries-Lequin de I, Kroesbergen HT, Visée WF, Aben DJM. Improving prevention in Dutch child healthcare in schools: development of a short questionnaire for early detection of psychosocial- and related child rearing problems. *Eur J Public Health* 1999;9(1999):200-4.

36   Bos C, Ede van J, Maarsingh EJ, Lako CL. De ontwikkeling van de KIVPA: de Korte Indicatieve Vragenlijst voor Psychosociale Problematiek bij Adolescenten. *Tijdschrift voor Jeugdgezondheidszorg* 30[56], 61. 1998.

37   Reijneveld SA, Vogels AGC. Vragenlijsten voor psychosociale problemen in de praktijk van de jeugdgezondheidszorg. In: Bonnet-Breusers AJM, Hirasing RA, Hoppenbrouwers K., Rensen H.B.H., Wagenaar-Fischer M.M., editors. *Praktijkboek Jeugdgezondheidszorg.* Elsevier; 2005. p. I 2.5-1-I 2.5-15.

38   Wilson JMG, Jungner G. *Principles an Practice of screening for disease.* Geneva: World Health Organization; 1968.

39   National Screening Committee. *First Report of the National Screening Committee.* London: Health Departments of the United Kingdom; 1998.

40   National Screening Committee. *Second Report of the UK National Screening Committee.* London: Health Departments of the United Kingdom; 2000.

41 Elliman DA, Dezateux C, Bedford HE. Newborn and childhood screening programmes: criteria, evidence, and current policy. *Arch Dis Child* 2002 July;87(1):6-9.

42 Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001 April;155(4):462-9.

43 Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001 April;155(4):462-9.

44 Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems among toddlers in Dutch preventive Child Healthcare. *Arch Pediatr Adolesc Med* 2004;158(august 2004):811-7.

45 Verhulst FC, Dekker MC, Van der EJ. Parent, teacher and self-reports as predictors of signs of disturbance in adolescents: whose information carries the most weight? *Acta Psychiatr Scand* 1997 July;96(1):75-81.

46 Goodman R, Ford T, Corbin T, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *Eur Child Adolesc Psychiatry* 2004;13 Suppl 2:II25-II31.

47 Wiefferink CH, Reijneveld SA, de Wijs J, Swagerman M, Campman D, Paulussen TG. Screening for psychosocial problems in 5-6-year olds: a randomised controlled trial of routine health assessments. *Patient Educ Couns* 2006 January;60(1):57-65.

48 Standards AaASD. Evaluation Of The Starting Together Initiative: Year Two 2005-06. San Diego: San Diego Unified School District; 2006.

49 Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997 July;38(5):581-6.

50 Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *Eur Child Adolesc Psychiatry* 1998 September;7(3):125-30.

51 Goedhart A, Treffers F, Widenfelt van B. Inquiring about mental health problems among children and adolescents: de Strenghts and Difficulties Questionnaire (SDQ) (in Dutch). *Maandbl Geestelijke Volksgezondh* 2003;58:1018-35.

52 Widenfelt van BM, Goedhart AW, Treffers PD, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003 December;12(6):281-9.

53 Embretson SE, Reise SP. *Item Response Theory for psychologists*. London: Lawrence Elbaum Associate Publishers; 20060.

2 Differences between professionals when identifying children with problems were not explained by actual problems present

A.G.C. Vogels

G.W. Jacobusse

F. Hoekstra

E. Brugman

M.R. Crone

S.A. Reijneveld

Abstract

Objective

To assess whether differences between individual Preventive Child Healthcare (PCH) professionals in the percentage of children they identify as having problems are larger than expected based on chance and whether such differences can be explained by differences in parent-reported problems or risk indicators.

Study Design

We used data from three community based studies among 3170 Dutch children aged eight to twelve, invited for a routine health examination. Parents filled in the Child Behavior Checklist (CBCL) and questions on demographic characteristics. After the examination PCH professionals registered whether they had identified any psychosocial problem in the child.

We examined differences between professionals in the rate of identified children. We used multilevel logistic regression analysis to assess whether such differences could be explained by parent reported problems or risk indicators.

Results

Significant differences between PCH professionals were found ($p < 0.001$). These differences could not be explained by parent-reported problems or risk indicators. The differences had the largest effect for children with a score above the CBCL clinical cut-off point.

Conclusion

Some PCH professionals are more likely to identify psychosocial problems than others, independently from parent-reported problems or other risk indicators.

2.1 Introduction

Emotional and behavioral problems are quite common among children, in all age groups. Such problems can lead to serious limitations in daily functioning and are likely to persist in later life.[1,2] However, only a minority of these children is treated for these problems. Verhulst et al. found that only 13% of children with behavioral or emotional problems were referred to mental healthcare.[3]

Community pediatric services, like those in the USA and the Netherlands, offering routine healthcare services to the population as a whole, are in a unique position to identify children with psychosocial problems. In the Netherlands this early identification is an explicit task of the existing Preventive Child Healthcare (PCH). Several studies have shown, however, that identification of children with psychosocial problems in community and primary pediatric care is less than satisfactory.[4-7] Clearly, identification of such problems is not always easy, but when done carefully, it can help to improve the prognosis of the children involved.[8]

Impressions from clinical practice lead us to believe that individual PCH professionals do not always share the same criteria when rating children's health. Inspection of data collected in other studies (e.g. Brugman et al.[9]) had revealed large differences between individual PCH professionals with respect to the proportion of children they identified as being (rather) unhealthy. In 2005 Crone et al. found significant differences between some PCH centers in the percentage of children identified as having problems.[10] They suggest that differences in protocols may be relevant in this respect, but also indicate that this can only partially explain the differences they found. In this study we want to assess a) whether there are systematic differences between individual PCH professionals in the proportion of children they identify as having psychosocial problems, b) whether such differences may be explained by differences in the prevalence of problems or differences in possible risk indicators among the groups of children seen by individual PCH professionals, and c) how large such differences are in relation to children's Child Behavior Checklist Total Problem Score (TPS).

2.2 Methods

2.2.1 Sample

We analyzed data from three previous studies that used similar methodologies with regard to sampling and data collection and the same or very similar measures and included children in the age range between eight and twelve years. The three studies used a similar two-step data collection procedure. First, we randomly sampled a number of regional PCH centers, after stratification by region and degree of urbanization of their district. Second, each PCH center provided a sample of children from those who were to be invited for a routine preventive health assessment. Study 1 assessed the prevalence of psychosocial problems in a national representative sample in 1997/1998,[7,11] in which 10,054 parents of children aged zero to fifteen participated (response rate: 93%). Study 2 examined the usability of three different questionnaires for the early detection of

psychosocial problems by PCH[12] among 2041 parents of children aged eight to twelve (response rate: 84%). Study 3 studied psychosocial problems in a national sample in 2004,[13] with 4776 parents of children aged fourteen months to twelve years participating (response rate: 85%). More detailed information about the data collection can be found elsewhere.[12-14] Differences between respondents and non-respondents by sex, age and ethnicity in the three studies were small, with Cohen's W effect size index varying between 0.01 and 0.11. The largest differences were found for ethnicity.

2.2.2 Procedure and measures

The data were collected in a standardized way as part of the routine preventive health assessment for which all Dutch children are invited regularly. Children were examined by those PCH professionals that were scheduled to see the children for the routine health examination. All participating PCH professionals received on site training by the researchers regarding protocol and the categorization of psychosocial problems. Data were collected by means of paper and pencil questionnaires for parents and for the PCH professional doing the routine preventive health examination. The questionnaires for parents were mailed to them along with the standard invitation for a preventive health assessment. After completion by the parents the questionnaires were returned to PCH, which sent them, unopened, to the researchers. During the preventive health assessment, the PCH professional obtained additional information from the parents on socio-demographic background, family characteristics and mental health history. Finally, after the health assessment, the PCH professional answered additional questions regarding his or her assessment of the child's psychosocial problems.

The dependent variables in this study were whether the PCH professional identified the child as having any psychosocial problem and whether he or she identified any problem for which the child was referred to mental healthcare or youth care. This information was obtained using the PCH questionnaire. The PCH professional was asked: "Does the child have a psychosocial problem at this moment?" (yes or no). Psychosocial problems were defined as emotional and behavioral problems of the child itself. Children with only risk indicators for the development of psychosocial problems had to be coded as having no problems. PCH professionals were explicitly asked to indicate all problems, whether light, moderate or severe. They also reported whether the problems identified lead them to refer the child to mental healthcare or to youth care.

PCH also asked parents whether the child had ever been treated because of psychosocial problems and if the child was currently under treatment. If so, he asked for the institution by which the child had been treated.

In each of the three studies, the parents or PCH provided information about the following possible risk indicators: number of (biological or non-biological) parents present in the family (one or two), parental employment status (1 or 2 fulltime or two part time jobs, one part time job or jobless), parental educational level (father's, if available, otherwise mother's: low (up to eight years in education), medium, (nine to twelve years in education) or high (more years in education)), age and gender.

The data sets of each study contained an identification code for the PCH professional who examined the child. One of the PCH professionals involved had participated in two

studies. We ensured that this professional's identification code was identical across the two studies. Finally, we knew the discipline of the PCH profession, either nurse or a physician.

Each study used the Child Behavior Checklist (CBCL) as a measure for the presence of emotional and behavioral problems. The CBCL was part of the questionnaire answered by parents. The 120 item CBCL was originally developed by Achenbach[15] and has been shown to be reliable and to have substantive concurrent and predictive validity.[1,16,17] In study 1 and 2, the Dutch 1991 version of the CBCL[18] was used. In study 3, the Dutch version of the 2001 American version of the CBCL[19] was used, which has been revised regarding the wording of some items and the construction of some of the subscales. Both versions allow for the calculation of a score on the CBCL Total Problem Scale (TPS), which includes information from all items and uses all items. We used the standard clinical cut-off points, as defined in the manuals.[18,19] in order to distinguish between children with and without problems. For the purpose of this study, the TPS scores of both versions are assumed to measure a very similar concept and therefore, essentially, to be identical. This assumption was checked in the analyses.

2.2.3 Analyses

We limited the analyses to children in the age range from eight to twelve years, of Dutch origin and currently not under treatment for psychosocial problems. The age range of eight to twelve was chosen because all three studies covered this age group: study 1 contained 1231 children in this age group, study 2 1825 and study 3 1351. Children from non-Dutch origin – i.e. at least one parent born outside the Netherlands – were excluded because of the following reasons. As people from non-Dutch immigrants tend to live in specific neighborhoods and tend to attend specific schools, some PCH professionals see far more children from non-Dutch origin than others. In our sample 67% of these children were examined by only 22% of the PCH professionals involved. Reijneveld et al. showed that problem identification by PCH among Dutch children was clearly associated with parent-reported problems; for children from non-Dutch origin such an association was not found.[20] This suggests strongly that problem identification among these groups is a more complicated process and should be studied separately from that among problem identification in general.

We also excluded children currently under treatment, because this is an almost certain indication of the presence of psychosocial problems, and treatment status was known to the PCH professionals. Finally, children with any missing data on the variables used in this study were excluded. The resulting sample, available for analysis, consisted of 3140 cases, 949 from study 1, 1444 from study 2 and 747 from study 3.

First, we assessed whether the differences between individual PCH professionals in the percentages of children identified as having problems were larger than might be expected on the basis of a simple binomial distribution. A simple chi-squared test could not be used because some physicians did only see a very limited number of children (in 7 cases, only 1 child). Therefore, we used a Monte Carlo procedure to attain an unbiased estimate of the exact significance level.

Secondly, we assessed to what extent these differences could be explained by differences in TPSs or the potential risk indicators mentioned before. We used bivariate and multivariate logistic regression analyses to assess the statistical significance of the relation of these characteristics and the discipline of the PCH professional – by themselves and in combination – with the identification of problems by PCH. The variables assessed this way are presented in Table 1. Those factors that showed a significant relationship were included in the multilevel analyses that followed as risk indicators. Then we performed a series of multilevel logistic regression analyses using three models. The first model only calculated the effect of the identity of the PCH professional as a level 2 factor on the likelihood of children being identified as having problems. In Model 2, the CBCL TPS score was added to the analyses. We plotted the logistic predictions based on the CBCL together with the empirical percentage of children identified by PCH. As the logistic prediction fitted the data very well, we decided to use the continuous CBCL score as a predictor in the model. In model 3 risk indicators were added to the analysis as level 1 factors. This way we assessed whether the effect of the PCH professionals' identity remained significant when TPS score and background characteristics were added to the equation. We used the Random Intercept Variance coefficient (RIV), a coefficient calculated in multilevel logistic regression analysis, to assess whether there was an effect of the level 2 professionals' identity. When the 95% Confidence Interval around this RIV lies above 0, the second-level effect may be considered significant. We estimated only random intercepts and no random slopes for predictors. These analyses were done using both the identification of any problem and the identification of problems for which children were referred as dependent variable. The multilevel analyses were done in S-Plus.[21] For parameter estimation we used Penalized Quasi Likelihood with first order Taylor linearization.

Some PCH professionals examined very few children. In order to determine whether the number of children examined had an effect on the results, we repeated all analyses including only professionals who had examined at least 20 children. To check for the assumption that the two different versions of the CBCL used in this study could be considered to be identical, we also repeated the analyses with CBCL version number added as a predictor and checked whether version number showed a significant effect.

As mentioned before, we removed children from non-Dutch origin from the analyses. It may be, however, that those professionals who see many of those children, tend to develop a different frame of reference and thereby contribute disproportionately to inter-individual variance. We checked for this by an additional multilevel analysis adding the percentage of non-Dutch children examined by individual PCH professionals (ranging from 0% to 78%) as a professional's characteristic to the model 3.

Finally, in order to gain some insight in the magnitude of the effects we found, we first estimated the relative size of the child-adjusted differences in identification between professionals using the Variance Partition Component (VPC).[22] The VPC indicates which part of the differences can explained by the level-2 clustering, i.e. inter-professional variability, for each value of the other variables included in the model. For these analyses we used two simplified models: one with the TPS score and gender as factor in the model, the other one with the TPS score and past treatment for psychosocial problems as predictors. Second, we calculated the model-predicted

24

probabilities of identification, in relation to TPS score. We did this for professionals on the 5th, 25th 50th 75th and 95th percentile. The probabilities were calculated using a simplified model with TPS as the sole predictor.

Table 1   *Child and family characteristics and CBCL Total Problems Scale score and properties of children identified as having problems*

	N	(%)	Of which identified as having problems
Total	3140	(100%)	20.7%
Age in years			
8	378	(12.0%)	31.0%
9	562	(17.9%)	21.7%
10	862	(27.5%)	18.4%
11	1022	(32.5%)	17.6%
12	316	(10.1%)	22.8%
Mean age in years (sd)	10.1	(1.2)	
Gender			
boys	1526	(48.6%)	24.2%
girls	1614	(51.4%)	17.4%
No of parents			
1	239	(7.6%)	39.3%
2	2901	(92.4%	19.2%
Job status			
1 or 2 full time / 2 part time jobs	2899	(92.3%)	19.7%
1 part time job	156	(5%)	33.3%
jobless	85	(2.7%)	31.8%
Highest educational level completed by parents			
primary level, up to 8 years	95	(3%)	33.7%
secondary level, from 9 to 14 years	1999	(63.7%)	22.0%
academic level, more than 14 years	1046	(33.3%)	17.1%
Ever treated for psychosocial problems			
No	2673	(85.1%)	17.8%
Yes	467	(14.9%)	37.5%
Examined by a			
physician	567	(18.1%)	18.3 %
nurse	2573	(81.9%	21.2%
Mean TPS score (sd)	17.0	(14.0)	

2.3 Results

Table 1 presents information about the sample used in this study. The 3140 children involved had been examined by 117 different PCH professionals; 61 of the PCH professionals had examined 20 children or more. The number of children examined by each individual PCH professional varied between 1 and 163, with a mean of 26.8.

About one fifth of the children was identified by the PCH professionals as having some psychosocial problem. Figure 1 shows that this proportion varied widely between individual professionals. Only PCH professionals who saw at least 10 children are incorporated in Figure 1.The black bars present the observed data. The grey bars present the expected numbers, based on the binomial distribution. Variation in the proportion of

identified children between PCH professionals was larger than could be expected based on fluctuation (p-value based on 10.000 Monte Carlo samples <0.001).

Figure 1   *Number of PCH professionals by proportion of children identified as having a problem, observed frequencies and expected frequencies based on the binomial distribution.*
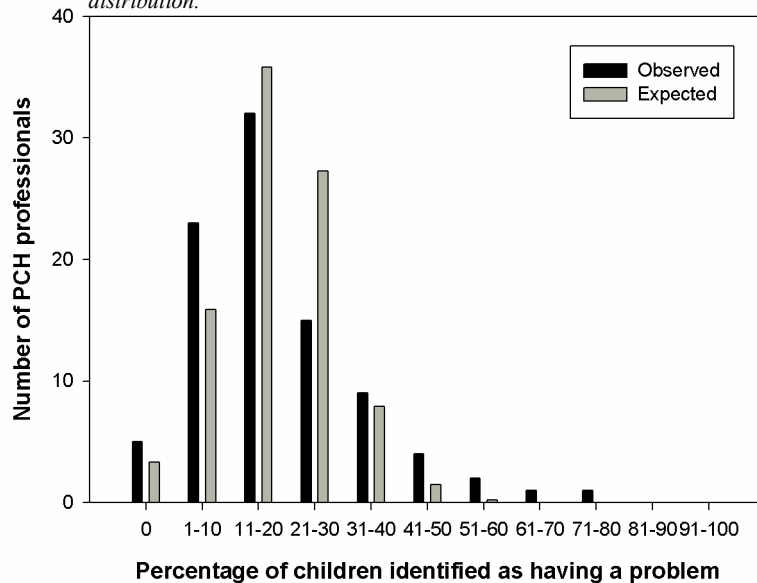
The percentage of children identified as having a psychosocial problem for which they were referred to mental healthcare or to youth care was 6%. The proportions of PCH professionals that referred none, some but less than 6%, 6 to 10% and 10% or more were 40%, 23%. 15% and 22%, respectively. Again, the variation between PCH professionals was larger than could be expected based on random fluctuation (p-value based on 10.000 Monte Carlo samples < .001).

In bivariate and multivariate logistic regression analyses most variables presented in Table 1 showed significant relationships with problem identification by PCH. However, we found no such a relationship for job status and being examined by a physician or nurse. So, age, gender, number of parents, educational level and treatment status were included in the multilevel analyses.

Table 2 shows the results of the multilevel analyses. Model 1 (first column) is the model with only variation between children an between professionals. The associated RIV was 0.34 with a confidence interval well above zero, indicating a statistically significant variation between professionals. The results of the model 2 (second column) showed that the TPS score was indeed related to identification by PCH professionals with statistical significance, but the second-level RIV remained statistically significantly higher than zero. The results of model 3 (third column) showed statistically significant effects of the risk indicators too, but again the variation between professionals remained statistically significant.

Table 2   *Results of the multilevel analyses on the identification of any problem in three models*

	Model 1		Model 2		Model 3	
		95% CI		95% CI		95% CI
Intercept	-1.36	-1.51 – -1.21	-2.35	-2.57 – -2.14	-1.01	-2.34 – .31
Random Intercept Variance (level 2)	0.34	0.20 – 0.57	0.43	0.26 – 0.71	0.39	0.24 – 0.66
			B	**95% CI**	**B**	**95% CI**
TPS score			0.05	0.05 – 0.06	0.05	0.04 – 0.05
Age					-0.19	-0.30 – -0.08
Gender					-0.35	-0.55 – -0.16
no of parents					0.87	0.56 – 1.18
past treatment					0.78	0.53 – 1.04
educational level 1					-0.63	-1.13 – -0.13
educational level 2					-0.85	-1.38 – -0.33
no of PCH professionals			117			
no of children			3140			
mean no (range) of children for each PCH profession:			26.8 (1-163)			

We repeated these analyses with the identification of problems for which children were referred as the dependent variable. In each of the models the RIV we found was significant: in the first model it was 0.75 (0.55 – 1.00); in the second model it was 0.83 ((0.62 – 1.11) and in the final model it was 0.83 (0.63 – 1.12).

The analyses which added CBCL version number as a predictor in the model, did not show a statistically significant effect for version number, nor did the second-level RIV associated with the professionals' identity change substantially. The results of the analyses in which the percentage of non-Dutch children examined by individual PCH professionals was added to model 3, did not lead to other conclusions, either The same applies for the analyses including only physicians who saw 20 children or more.

Figure 2 presents information about the size of the differences between professionals, in relation to the TPS score and the child's gender. The size of the differences is clearly related to the TPS score. For boys the differences between individual PCH professionals are most marked when they had a TPS score of about 40. This is near the CBCL clinical cut-off point (36-40, depending on gender and age). For boys with a CBCL TPS around 40 about 9% of the variance in problem identification must be attributed to the identity of the PCH professional by whom they are examined. For girls the differences are most marked in relation to a TPS score of about 50, well above the standard TPS cut-off point.

Figure 3 presents the effect sizes in relation, again, to the TPS and stratified to whether children underwent any treatment for psychosocial problems in the past. The relation to the TPS was similar to that presented in Figure 2. For children who had been treated, individual differences between professionals were more marked at somewhat higher TPS scores than for those who had never been treated.

Figure 2   *Degree of inter-professional variability, as measured by the Variance Partition*
              Components, in relation to TPS score and gender
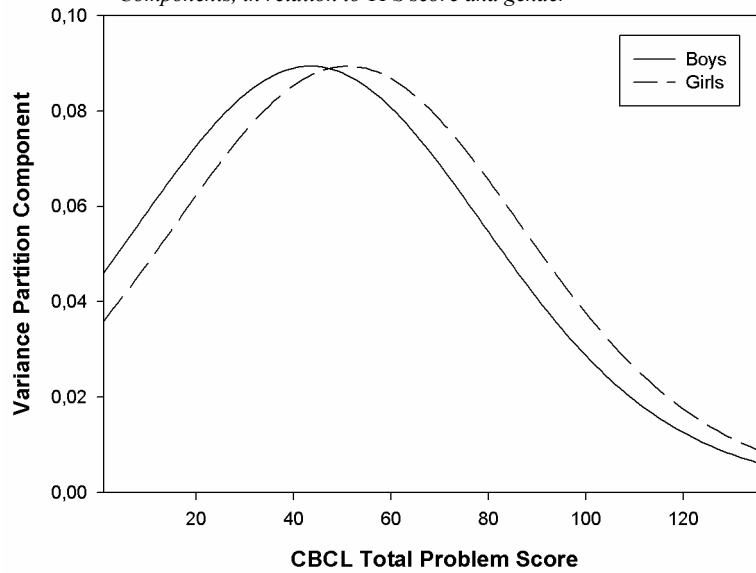
Figure 3   *Degree of inter-professional variability, as measured by the Variance Partition*
              Components, in relation to TPS score and having been treated for psychosocial
              problems in the past
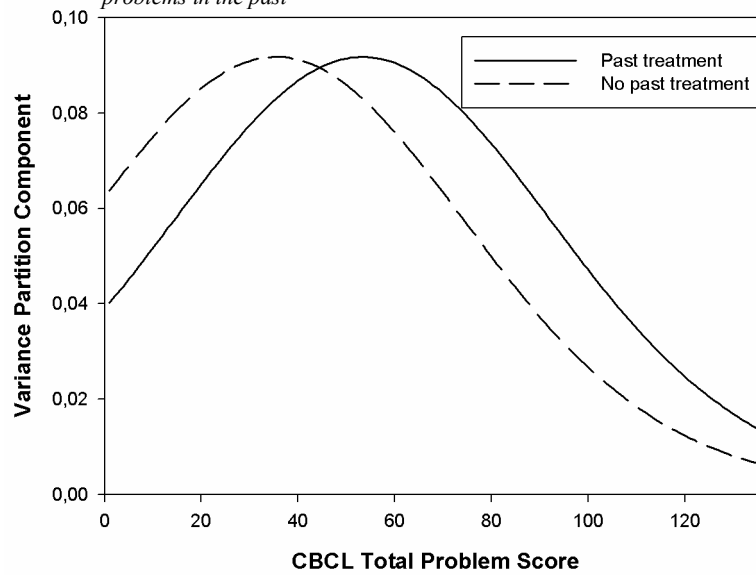
Figure 4   *Model predicted probabilities that PCH professionals identify a child as having a problem in relation to a TPS score, for PCH professionals on the 5th , 25th , 50th, 75th and 95th percentile*
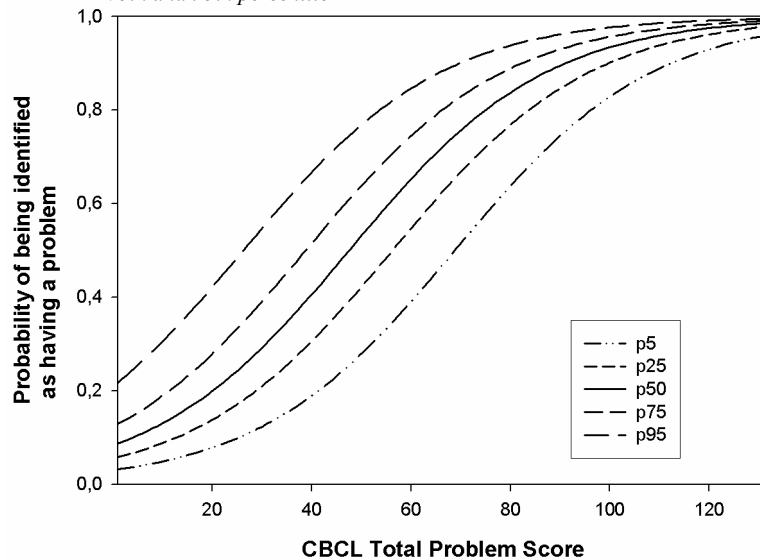
Figure 4, finally, presents the model-predicted probabilities of problem identification in relation to TPS score. The solid line in the center indicates the mean probability, indicating that, on average, children with a TPS of 40 have a probability of about 40% of being identified as having problems. For 50% of all professionals, those between the two lines indicating the 25th and 75th percentiles, this probability lies between about 30 and 50%. For 5% of the professionals this chance is less than 20% and also for 5% this probability is more than 65%. Again we see that the differences at very low or very high TPS scores are much smaller. Yet, on a large part of the scale, the differences remain substantial.

2.4 Discussion

Main results

This study showed that there are important differences between individual PCH professionals in the identification of children with psychosocial problems, and that these differences cannot be explained by parent-reported problems on the CBCL or by differences in the child's background characteristics. In other words, some PCH professionals tend to perceive more children as having problems than other professionals. This applies both to the identification of any problem and to the identification of problems for which children were referred to mental healthcare or youth care. Differences in problem identification between professionals were most marked for children with a TPS score above the clinical cut-off point, especially for girls – as compared to boys – and for children who had been treated for psychosocial problems in the past, compared to children who had never been treated.

29

Limitations

The three studies used data that are representative for all children of Dutch origin in this age group in the Netherlands under care by PCH. All three studies were done in the context of regular care in Dutch PCH. Yet, there are reasons to suppose that the differences we found are an underestimation of the effects in standard practice. First, we excluded children from non-western origin. We previously showed that identifying problems among children from non-Dutch origin probably is more difficult than among children from Dutch origin, due *e.g.* to language problems, cultural barriers and so on.[20] We expect, therefore, that differences between individual PCH professionals have an even greater effect when children from non-Dutch origin are involved. Furthermore, although data collection was done in the context of regular care, PCH professionals who participated in the studies, may be – on average – more interested in the identification of children with psychosocial problems. If this is true, the variation between PCH professionals in the field will probably be larger that the variation in our sample.

The available data sets posed some limitations to the variables we could use in the analyses. For example, some PCH professionals work mainly in socially deprived neighborhoods. Such work setting characteristics may evidently be related to differences in the proportion of children identified as having problems.[23,24] We could not include such factors in the analysis. However, at least a part of the effect of this factor will be related to a higher problem level in problem areas and this will be reflected in a higher TPS of the children involved. As far as this is the case, in our analyses the effects of this factor will be attributed to TPS. Therefore, in our view, it seems unlikely that this factor could explain all of the systematic inter-individual variance we found, and we believe, despite the limitations of this study, that our results indicate actual differences between PCH professionals in the Netherlands. Whether our conclusions may be generalized to other settings is difficult to answer, as we found no other studies assessing this problem as we did. However, it would be interesting to investigate this issue in settings, like community-based pediatric centers and primary healthcare, in which staff with limited psychological and psychiatric training has to identify children or adults with problems.

It would have been very interesting to determine which characteristics of PCH professionals (e.g. experience and education) could explain the differences between them. Unfortunately, we did not have any data on PCH professionals characteristics.

Fit with the literature

To our knowledge this is the first study that tried to assess differences between individual healthcare providers in the identification of children with psychosocial problems, as it occurs in large-scale standard practice. Many studies[25-28] have reported on interrater reliability in the field of mental health assessment, but those studies used small samples and mostly concerned specific instruments. Such studies are less suited for assessing the effects of differences between individual healthcare providers in real life, as the study design itself introduces deviations from standard care: two or more persons assessing the same child, using video taping, observations through one-way screens, case descriptions or vignettes etc.

Implications

Are the differences that we found serious? For children with very low TPS scores the effects we found were relatively modest – and most children score in the lower TPS range. When a child has a TPS of about 40, the probability that it will be identified as having problems lies between 30% and 50% when seen by half of the PCH professionals. This is already a remarkable range. Yet, for 25% of all PCH professionals, the probability is still higher and also for 25% this probability is lower. In other words, the variation is largest at relatively high scores, just where the identification by PCH matters most. In our view, therefore, the differences that we found are serious, indeed. We must, however, take into account that the majority of the children have relatively low scores and that our model is therefore based mainly on those children. For that reason, our study should be replicated in a sample with more higher scoring children.

Other studies reported that PCH did not identify any problem in about half of all children with a TPS score above the clinical cut-off point.[7,10] This finding in itself does not necessarily point to under-identification of problems, as a clinical TPS score is only a valid and reliable indication for problems, not absolute and certain evidence.[18] The clinical assessment by the PCH professional might be more correct, e.g. because he or she identifies problems that are not or insufficiently indicated by the CBCL. However, if this assumption were the main explanation for the lack of agreement between PCH identification and CBCL TPS, we would expect it to be randomly distributed over individual professionals.

This study showed that this is not the case. Identification of emotional and behavioral problems by PCH should thus be improved, thereby diminishing inter-professional variation. The question then remains how to reduce the inter-professional variability. The literature offers several suggestions which can help to do so. Several studies showed that good quality questionnaires can improve problem identification by PCH.[8,12,29,30] Wiefferink et al. showed that training PCH professionals and the use of detailed protocols led to an increase in the sensitivity and specificity of problem identification by PCH and thus to a reduction in inter-professional variability. The effects in their study diminished in time, though.[31] This shows that attention should be paid to continuation of the method and protocols used. Crone et al. studied the sustainability of a health education program in Well-Baby clinics.[32] They found that factors such as perceived self-efficacy, responsibility, training attendance, participation in the adoption decision and level of institutionalization were related to the degree to which physicians and nurses used the program. Such findings may be very relevant for protocols for the identification of psychosocial problems, too. We are not aware of such implementation studies in this field.

The results of this and other studies[14,33] clearly indicate that the development of high-quality protocols and tools and an effective implementation program with continuous attention to adherence to the protocol are highly desirable. Moreover, the time for applying tools during routine examinations may have to be increased, too. Further research has to show how this could be achieved most effectively.

Conclusion

There are systematic differences between individual PCH professionals in the number of children they identify as having problems, and these differences cannot be explained by problems – as indicated by a clinical CBCL TPS – among these children or other risk indicators.

Clearly, PCH professionals need to improve the identification of problems among children they examine, especially as other studies have shown that such an improvement is possible.[8,29-31]

The inter-individual variability that we found shows that there are large opportunities to improve the identification of psychosocial problems among children. In order to do so, high-quality protocols, including the necessary tools, an effective and structural implementation program and sustained attention to adherence to protocols have to be developed.

References

1   Verhulst FC, Dekker MC, Van der EJ. Parent, teacher and self-reports as predictors of signs of disturbance in adolescents: whose information carries the most weight? *Acta Psychiatr Scand* 1997 July;96(1):75-81.

2   Moffitt TE, Caspi A. Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Dev Psychopathol* 2001;(13):355-75.

3   Verhulst FC, Ende van der J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997 July;36(7):901-9.

4   Glascoe FP, Dworkin PH. The role of parents in the detection of developmental and behavioral problems. *Pediatrics* 1995 June;95(6):829-36.

5   Glascoe FP. Screening for developmental and behavioral problems. *Ment Retard Dev Disabil Res Rev* 2005;11(3):173-9.

6   Lavigne JV, Binns HJ, Christoffel KK et al. Behavioral and emotional problems among preschool children in pediatric primary care: prevalence and pediatricians' recognition. Pediatric Practice Research Group. *Pediatrics* 1993 March;91(3):649-55.

7   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001 April;155(4):462-9.

8   Durlak JA, Wells AM. Primary prevention mental health programs for children and adolescents: a meta-analytic review. *Am J Community Psychol* 1997 April;25(2):115-52.

9   Brugman E, Meulmeester JF, Spee-van der Wekke J, Beuker RJ, Radder JJ. *Peilingen in de jeugdgezondheidszorg*. PGO-Peiling 1993/1994. Leiden: TNO Preventie en Gezondheid; 1995.

10   Crone MR, Wiefferink CH, Schuller AA, Reijneveld SA. *Ontwikkeling, gezinssituatie, vrijetijdsbesteding van kinderen van 0-12 jaar. Provincie Zuid Holland*. Leiden: TNO Kwaliteit van Leven; 2005.

11   Brugman E, Reijneveld SA, Hollander-Gijsman ME, Burgmeijer RJF, Radder JJ. *Peilingen in de Jeugdgezondheidszorg PGO Peiling 1997/1998: melkvoeding en huilgedrag van zuigelingen, psychosociale gezondheid van 1-15 jarigen*. Leiden: TNO Preventie en Gezondheid; 1999.

12   Vogels AGC, Crone MR, Hoekstra F., Reijneveld SA. *Drie vragenlijsten voor het opsporen van psychosociale problemen bij kinderen van zeven tot twaalf jaar.* Leiden: TNO Kwaliteit van leven; 2005.

13   Zeijl E, Crone MR, Wiefferink CH, Keizenkamp S, Reijneveld SA. *Children in the Netherlands [in Dutch]*. The Hague: SCP; 2005.

14   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001 April;155(4):462-9.

15   Achenbach TM, McConaughy SH, Howell CT. Child/Adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletin* 1987;101(2):213-32.

16   Ferdinand RF, Visser JH, Hoogerheide KN et al. Improving estimation of the prognosis of childhood psychopathology; combination of DSM-III-R/DISC diagnoses and CBCL scores. *J Child Psychol Psychiatry* 2004 March;45(3):599-608.

17   Koot HM, van den Oord EJ, Verhulst FC, Boomsma DI. Behavioral and emotional problems in young preschoolers: cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *J Abnorm Child Psychol* 1997 June;25(3):183-96.

18   Verhulst FC, Koot HM. *Manual for the CBCL/4-18 [in Dutch]*. Rotterdam: EUR/Sophia Kinderziekenhuis; 1996.

19   Achenbach TM, Rescorla LA. *Manual for the ASEBA School-Age Forms & Profiles*. Burlington VT: University of Vermont, Research Center for Children, Youth and Families; 2001.

20   Reijneveld SA, Harland P, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Psychosocial problems among immigrant and non-immigrant children--ethnicity plays a role in their occurrence and identification. *Eur Child Adolesc Psychiatry* 2005 May;14(3):145-52.

21   Venables WN, Ripley BD. *Modern Applied Statistics with S-PLUS*. 3 ed. New York: Springer; 1999.

22   Goldstein H, Browne WJ, Rasbash J. Partitioning variation in multilevel models. *Understanding Statistics* 2002;(1):223-32.

23   Harland P, Reijneveld SA, Brugman E, Verloove-Vanhorick SP, Verhulst FC. Family factors and life events as risk factors for behavioural and emotional problems in children. *Eur Child Adolesc Psychiatry* 2002 August;11(4):176-84.

24   Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Area deprivation and child psychosocial problems--a national cross-sectional study among school-aged children. *Soc Psychiatry Psychiatr Epidemiol* 2005 January;40(1):18-23.

25   Dyrborg J, Larsen FW, Nielsen S, Byman J, Nielsen BB, Gautre-Delay F. The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice--substance and reliability as judged by intraclass correlations. *Eur Child Adolesc Psychiatry* 2000 September;9(3):195-201.

26   Bird HR, Canino G, Rubio-Stipec M, Ribera JC. Further measures of the psychometric properties of the Children's Global Assessment Scale. *Arch Gen Psychiatry* 1987 September;44(9):821-4.

27   Esbensen AJ, Rojahn J, Aman MG, Ruedrich S. Reliability and validity of an assessment instrument for anxiety, depression, and mood among individuals with mental retardation. *J Autism Dev Disord* 2003 December;33(6):617-29.

28   Veloski J, Tai S, Evans AS, Nash DB. Clinical vignette-based surveys: a tool for assessing physician practice variation. *Am J Med Qual* 2005 May;20(3):151-7.

29   Reijneveld SA, Vogels AG, Brugman E, van Ede J, Verhulst FC, Verloove-Vanhorick SP. Early detection of psychosocial problems in adolescents: how useful is the Dutch short indicative questionnaire (KIVPA)? *Eur J Public Health* 2003 June;13(2):152-9.

30   Vogels T, Reijneveld SA, Brugman E, Hollander-Gijsman M, Verhulst FC, Verloove-Vanhorick SP. Detecting psychosocial problems among 5-6-year-old children in preventive Child Healthcare: the validity of a short questionnaire used in an assessment procedure for detecting psychosocial problems among children. *Eur J Public Health* 2003 December;13(4):353-60.

31   Wiefferink CH, Reijneveld SA, de Wijs J, Swagerman M, Campman D, Paulussen TG. Screening for psychosocial problems in 5-6-year olds: a randomised controlled trial of routine health assessments. *Patient Educ Couns* 2006 January;60(1):57-65.

32   Crone MR, Verlaan M, Willemsen MC et al. Sustainability of the prevention of passive infant smoking within well-baby clinics. *Health Educ Behav* 2006 April;33(2):178-96.

33   Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems among toddlers in Dutch preventive Child Healthcare. *Arch Pediatr Adolesc Med* 2004;158(august 2004):811-7.

3 Early detection of psychosocial problems in adolescents.
How useful is the Dutch short indicative questionnaire (KIVPA)?

S.A. Reijneveld

A.G.C. Vogels

E. Brugman

J. van Ede

F.C. Verhulst

S.P. Verloove-Vanhorick

Abstract

Background:

Psychosocial problems, such as behavioural, emotional, and educational problems, are highly prevalent among children and adolescents. Early treatment may reduce these problems, if accurately identified. Validated questionnaires may support identification. The aim of this study is to assess the psychometric qualities of such a questionnaire, the Short Indicative Questionnaire for Psychosocial problems among Adolescents (KIVPA), and to determine whether it is suitable for and adds to the early detection of psychosocial problems among adolescents.

Methods:

Data came from a national sample of 1,440 Dutch adolescents, using the KIVPA, the Child Behavior Checklist (CBCL), and the Youth Self Report (YSR). Of these, 1,248 provided data on all questionnaires (77.8%). The scale structure of the KIVPA was assessed; its sensitivity and specificity using CBCL, YSR and referral for psychosocial problems as criteria; and its contribution to detecting CBCL and YSR problems.

Results:

The KIVPA is mostly uni-dimensional but the variance explained by its main factor is relatively low. The Total KIVPA score discriminates between adolescents with and without problems on the three criteria. Using a clinical YSR Total problem score as criterion, sensitivity and specificity are 0.82 and 0.85, respectively, at the proposed cut-off (area under the ROC curve: 0.92; 95% confidence interval (CI) 0.90 – 0.95). The odds ratio of a clinical YSR score for an elevated KIVPA score is 29.1 (95%-CI: 14.4-59.1), although the KIVPA mainly covers internalizing problems.

Conclusion:

The KIVPA has added value in the early detection of internalizing psychosocial problems, but is not sufficiently efficient.

3.1 Introduction

Psychosocial problems, such as behavioural, emotional, and educational problems, are highly prevalent among children and adolescents, and may severely interfere with everyday functioning. Only a minority of the children with such problems receive mental healthcare.[1-4] In a study conducted among more than 2,000 Dutch children,[4] only 13% of the children with behavioural and emotional problems had been referred to mental health services in the year before the assessment.

In the Netherlands, preventive child healthcare is one of the most important low-threshold services for the early detection of psychosocial problems in children.[5] This preventive healthcare is systematically offered to all children living in the Netherlands by community physicians and nurses working in preventive Child Healthcare services (Child Health Professionals, CHPs).[6,7] As part of this system, more than 90% of all children undergo three to four assessments by a CHP during their school careers, in both primary and secondary school.[8] At present, new legislation has been started to support this system and the identification of psychosocial problems as part of it.[5]

We previously reported on the degree to which Dutch CHPs identified and managed psychosocial problems in children aged 4-15 years.[9,10] One or more psychosocial problems were identified in 25% of all children, and one in five of the identified children were referred for further diagnosis and treatment. Results further showed that identification of psychosocial problems in children and subsequent referral were six times more likely in the 8% with serious parent-reported problem behaviour (measured by the Child Behavior Checklist,[11,12] a well-validated questionnaire for emotional and behavioural problems). However, CHPs identified no psychosocial problems in 43% of these children and therefore undertook no action. On the basis of this it was concluded that screening for psychosocial problems may be a promising option to reduce these problems, but that accuracy of the identification should be enhanced.[9,10]

One way to improve the early identification of mental health problems in children and adolescents may be the use of validated questionnaires. For instance, a meta-analysis by Durlak and Wells[13] shows that early treatment is much more effective if cases are identified in such a way. In the Netherlands, a number of questionnaires have been developed to this end in preventive child healthcare.[5,14,15] Such questionnaires should, of course, have good psychometric properties and should also be short: usually only ten minutes are available for the routine health assessment of a child, and mental health is only part of the assessment.

One of the questionnaires used is the Short Indicative Questionnaire for Psychosocial problems among Adolescents (abbreviated in Dutch as KIVPA).[14,16,17] The KIVPA aims at the detection of psychosocial problems in 12-18-year-old adolescents. Adolescents have to complete it in class or at home before their routine health assessment. The KIVPA has been developed by a working group of Dutch CHPs and epidemiologists. It consists of three parts, each with a different background: A) an indicative scale (for mental health; eight items); B) a psychosomatic scale (nine items);[18] C) a self-analysis scale (fourteen items). Answers on scale A and C are dichotomised (0/1), on scale B they are 0/1/2 coded. A summary score ranging from 0 to 21 is computed on the basis of a weighing of each part. Items and formula for this score are presented in Table 1.

A first validation study examined the psychometric quality of the questionnaire among 3,405 adolescents who were examined routinely. Their KIVPA scores were compared with those of 317 adolescents questioned during the intake for ambulatory mental healthcare.[14,17] Results show that mean scores of the first group were much lower than those of the second group. The authors propose a cut-off point of 6 and higher to be indicative of psychosocial problems. At this cut-off, 17% of the first group have an elevated score compared with 70% of the second group. In the first group, elevated scores occur far more often among girls than boys (23% vs. 10%). Further information on the total score is lacking, however, a fact that has been criticised.[19,20] Despite this, at least half of all regional Dutch departments for preventive child healthcare use the KIVPA at present (C. de Rover, written personal communication) and it is mentioned as such in the Dutch listing of Basic Tasks for preventive child healthcare.[5]

The aim of the present study is to assess the psychometric qualities of the KIVPA and whether it is suitable for and adds to the early detection of psychosocial problems among adolescents.

3.2 Methods

This study is based on a community sample of adolescents for whom data are available regarding the KIVPA, the CBCL, the Youth Self-Report (YSR) and the identification and management by a CHP.

3.2.1 Population
The sample was obtained using a two-stage selection procedure. In the first stage, a random sample of 19 of the 63 Child Healthcare Services was drawn, after stratification by region and degree of urbanization of their district. In the second stage, each Healthcare Service provided a sample of 75 children from the second grade of secondary school by inviting all children in three school classes of different levels to participate. Of the total sample of 1,604 eligible adolescents, 1,440 participated, 1,326 provided complete data on the KIVPA, and 1,248 on all questionnaires (89.8, 82.7, and 77.8% of the original sample, respectively; of all *participating adolescents*, 89.9% filled out the KIVPA). All three aforementioned groups were representative of the total sample, selective non-response was not found (on non-participating adolescents, data regarding gender, age, living area, ethnicity and mental health history were obtained from CHP files). Details have been presented elsewhere.[9,10] Analyses were restricted to those adolescents who provided data on all questionnaires, to make interpretation easier.

3.2.2 Data collection
The data were collected in a standardized way during routine preventive health assessments, from October 1997 to June 1998. The design of the study had been approved by the local Medical Ethical Committee. The KIVPA,[14] the YSR,[21,22] and the CBCL,[11,12] were mailed to adolescents, along with the standard invitation to the preventive health assessment. Adolescents completed the KIVPA and the YSR and put them in a sealed envelope; parents did the same with the CBCL. Adolescents gave both

envelopes to the CHP who passed them on to the researchers without opening them (whereas normally, the CHP would partially base the interview on the KIVPA). The CHP interviewed the child (and sometimes the parents; 11% were accompanied by a parent) regarding mental health and background and examined the child. After each assessment, the CHP filled out the following question: 'Does the child have a psychosocial problem, at this moment?' (yes, no), and scored its severity (mild, moderate or severe) and the type of the problem(s) identified, using a pre-coded list. Children who had only risk-indicators for the development of psychosocial problems, such as parents with psychiatric problems or other family problems, had to be coded as 'no'.

The YSR and the CBCL were used respectively to assess adolescent's and parent's report of the behavioural and emotional problems of the adolescent during the preceding six months. Both questionnaires are of a similar nature, but are worded differently. Their (good) reliability and validity have been established.[11,12,21,22] For this article we used only the problem items of both questionnaires and computed scores for nine syndrome subscales, two broad-band groups of syndromes designated Internalizing and Externalizing, and a Total Problem score. Regarding the Total Problem and broad-band scales, adolescents were also allocated to a normal range or a clinical range, using the 90th percentile of the Dutch normative sample as cut-off.[12,22]

3.2.3 Analysis

In the analysis the psychometric properties of the KIVPA and its added value in identifying psychosocial problems were assessed. Regarding psychometric properties, first the scale structure of the questionnaire was assessed using principal component analysis (PCA) and the internal consistency of each scale was computed. Regarding this, the approach proposed by the developers was followed. Next, the validity of the KIVPA was assessed by using dichotomized (normal vs. clinical) CBCL and YSR (Total Problem and Internalizing/Externalizing scales), and being referred because of psychosocial problems by the CHP, as criteria.[23]

Regarding the added value of the KIVPA in identifying psychosocial problems, the odds of identification of mental health problems (i.e. a clinical YSR and CBCL Total Problem score) by an elevated score on the KIVPA was assessed. This was repeated with adjustment for social and demographic risk indicators which are known to the CHP and might help him or her to identify psychosocial problems,[9,10] and for adolescents with and without CHP-identified problems. Regarding social and demographic risk indicators, adolescents with missing data were retained in the logistic regression models by creating separate dummies for the missing category of each variable. As this may lead to biased results,[23] all analyses were repeated, omitting all adolescents with missing data on these variables (remaining sample: 1,064 adolescents). Because the latter results were very similar, we do not present them here.

All analyses were made with SPSS 10.0 for Windows,[24] and were repeated for boys and girls separately. Results for these subgroups are given only when they differ significantly (p<0.05).

3.3 Results

Close to 20% of the sample had a score on the KIVPA of 6 and higher. In general, girls have higher scores on the KIVPA than boys, on the Total scale and all subscales (compare Table 2).

3.3.1 Structure of the questionnaire and reliability

The exploratory PCA on the Total KIVPA yielded eight components with eigenvalues higher than 1, of which one dominates (eigenvalues: 5.9, and ranging from 1.1 to 1.3, respectively). This implies that it mostly measures one construct. The proportion of the variance (i.e. the variation in answers) explained by this construct is rather low, however. The same holds for the loadings of some items (i.e. their association with this construct): see Table 1. Repetition of this analysis with a specified number of three components did not confirm the postulated three subscales, with either varimax rotation or with oblique rotation.

Next, the items from the three parts of the questionnaire were analysed separately (i.e. a PCA was performed on the items of the Indicative subscale, of the Psychosomatic subscale and of the Self-analysis subscale, respectively: see Table 1). This yielded one dominating principal component for the Psychosomatic subscale (eigenvalue: 2.6), but two for the Indicative subscale (eigenvalues 2.1 and 1.1), and four for the Self-analysis subscale, of which one dominates (eigenvalues: 3.3, and 1.3 to 1.0, respectively). The first principal components of the three subscales were also reasonably associated (correlation coefficients from 0.52 to 0.62), which indicates that they measure a rather similar construct. Measures of reliability were generally higher for girls than for boys.

Though the developers proposed a PCA on the dichotomized responses, all analyses were repeated on the original three-digit responses. This only marginally affected results regarding the factor structure, but results regarding reliability generally worsened (for the Indicative subscale, Cronbach's alpha dropped to 0.19).

3.3.2 Validity

The validity of the KIVPA was assessed using the CBCL, the YSR and the fact of being referred by the CHP because of psychosocial problems (without knowledge of the KIVPA score) as criteria. Mean KIVPA Total scores were higher for adolescents with a clinical Total Problem score on the CBCL and YSR, for all adolescents and for those not under treatment by mental health services. They were also higher for the referred group (Table 3).

Table 1   *Results of four principal component analyses, presenting the loadings of the items on the main principal component from analyses of (a) all items of the KIVPA,[a] (b) the items of only the Indicative subscale, (c) the items of only the Psychosomatic subscale, and (d) the items of only the Self-analysis subscale (n=1.248)[b]*

	(a)	(b)	(c)	(d)
Indicative subscale				
What do you think of your own health?	0.50	0.62		
How do you feel when you're at home?	0.50	0.66		
Do you find it easy to talk with your parents or caretakers easily?	0.47	0.65		
Do you worry about the future?	0.47	0.51		
Do you think that you have a sufficient number of friends?	0.40	0.41		
Are you happy about your appearance?	0.45	0.52		
Have you ever had a sexual experience with someone against your will?	0.16	0.25		
Have you recently felt so restless or agitated that you took a sedative or hypnotic because of that?	0.23	0.32		
Psychosomatic subscale				
Do you sometimes feel listless? (not feeling up to anything)	0.45		0.55	
Do you sometimes feel weary without knowing why?	0.51		0.66	
Do you sometimes have a headache because of stress?	0.45		0.54	
Does it happen sometimes that you do not feel like eating?	0.32		0.47	
Do you find it difficult to fall asleep?	0.44		0.54	
Do you have a sensitive 0073kin so that you easily get rash, spots or itch because of something?	0.26		0.34	
Do you sometimes have stomach-ache, around your navel?	0.42		0.55	
Do you sometimes feel that you cannot relax your muscles properly?	0.47		0.56	
Do you sometimes burst out crying although there is not much reason for it?	0.53		0.58	
Self-analysis subscale				
I feel good about myself [c]	0.33			0.34
I am rebellious or disobedient	0.30			0.27
I often feel unsure of myself	0.62			0.69
I worry a lot	0.64			0.68
I am independent [c]	0.20			0.28
I am often short-tempered or aggressive	0.33			0.34
I am often nervous or tense	0.51			0.55
I am a happy person [c]	0.38			0.39
I am close-mouthed, withdrawn	0.51			0.62
I often feel lonely	0.60			0.67
I am very shy	0.34			0.43
I am spontaneous [c]	0.21			0.26
I often feel down or depressed	0.65			0.68
I like to do a lot of things [c]	0.16			0.17
Internal consistency (Cronbach's alpha)	0.84	0.57	0.68	0.73
Percentage of total variance explained by first principal component	18.9	26.3	28.9	23.9

a: A KIVPA Total score is obtained by summing the scores of the three subscales according to the following formula (Indicative subscale + (sum score Psychosomatic subscale / 3) + (sum score Self-analysis subscale / 2)).

b: KIVPA = Short Indicative Questionnaire for Psychosocial problems among Adolescents.

c: Items are coded in reverse for computing the sum score.

Table 2   Descriptive statistics on the KIVPA Total scale and the KIVPA subscales, overall and by gender (n=1,248)

	Total	Indicative	Psychosomatic	Self-analysis
All (n=1,248)				
Mean (SD)	3.81 [a] (2.88)	1.05 [a] (1.29)	4.39 [a] (2.70)	2.49 [a] (2.38)
Median	3 [b]	1 [b]	4 [b]	2 [b]
Range	0-17	0-7	0-14	0-13
90th percentile value	8	3	8	6
Score 6 and higher	19.6% [c]	-	-	-
Boys (n=597)				
Mean (SD)	3.22 (2.48)	0.87 (1.13)	3.67 (2.40)	2.15 (2.16)
Median	3	0	3	2
Range	0-17	0-7	0-11	0-13
90th percentile value	7	2	7	5
Score 6 and higher	12.9%	-	-	-
Girls (n=651)				
Mean (SD)	4.36 (3.10)	1.22 (1.40)	5.05 (2.79)	2.80 (2.52)
Median	4	1	5	2
Range	0-17	0-7	0-14	0-13
90th percentile value	9	3	9	6
Score 6 and higher	25.7%	-	-	-

a: Statistically significant differences by gender (p<0.001), t-test.
b: statistically significant differences by gender (p<0.001), Mann-Whitney U-test.
c: statistically significant differences by gender (p<0.001), chi-square test.

Figure 1   Receiver Operating Characteristic curves for the Total KIVPA score using the YSR Total Problem, Internalizing and Externalizing scales and the CBCL Total Problem score as criteria
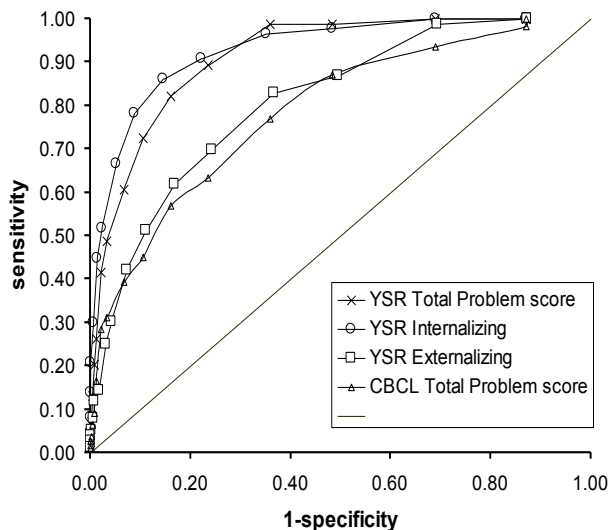
Subsequently, the degree to which the score on the KIVPA is indeed elevated in the case of psychosocial problems as measured by these three criteria (i.e. sensitivity), and the degree to which it is 'normal' in the case of absence of these problems (i.e. specificity) were assessed. Using the YSR as criterion, the KIVPA score is elevated for 82% of the adolescents with a clinical YSR Total Problem score. The reverse, a normal KIVPA score among adolescents with a normal YSR score, holds for 85% of the adolescents. The latter implies that 15% of the adolescents with a normal YSR score have an elevated KIVPA score. As most adolescents have a normal YSR score (in this sample about 92%), this low percentage still implies that only a minority (28%) of all adolescents with an elevated KIVPA score have a psychosocial problem as measured by the YSR. This is labeled in Table 3 as the positive predictive value (of an elevated KIVPA score). For the three criteria, the sensitivity of the KIVPA at a cut-off of 6 ranged from 0.44 to 0.86 and its specificity from 0.83 to 0.86. Lowest values concern Externalizing scales. Areas under the receiver operating characteristic (ROC) curves ranged from 0.68 to 0.93 (Table 3 and Figure 1).

To gain further insight into the contents of the KIVPA, its associations with scores on all YSR syndrome scales were assessed, as these come from the same informant (the adolescent). Resulting (Spearman) correlation coefficients were highest for the Anxious/Depressed and Withdrawn syndrome scales (0.68 and 0.59, respectively) and for Attention Problems and Identity Problems (only boys) (both 0.52). The first two of these are both part of the Internalizing broad-band of the YSR (for the third syndrome that is part of the Internalizing broad-band, Somatic Complaints, it was 0.49). Correlation coefficients for the other syndromes ranged from 0.46 to 0.37 (by decreasing value: Aggressive Behaviour, Social Problems, Thought Problems and Delinquent Behaviour), the first and the last constituting the Externalizing broad-band.

Finally, an elevated score on the KIVPA seems to add to the identification of psychosocial problems as measured by a clinical score on the CBCL and the YSR. In Table 4, odds ratios are presented which express the likelihood of a clinical score on these questionnaires if an adolescent has an elevated score on the KIVPA. Odds ratios are much higher, however, for the YSR, and for Internalizing Problems. Adjusting for background characteristics of the adolescent, which may help the CHP in identifying psychosocial problems, increased some of the odds ratios. This indicates that the KIVPA indeed provides additional information that is helpful for identification. Furthermore, some odds ratios were higher for adolescents in which the CHP, without knowledge of the KIVPA score, had not identified problems. This suggests that the KIVPA may, in particular, support the identification of problems not otherwise identified.

Table 3   Test characteristics of the KIVPA, using the CBCL, the YSR and being referred because of psychosocial problems, as criteria. Rank correlation coefficients, and sensitivity, specificity, positive and negative predictive value, and area under the receiver operating characteristic (ROC) curve at a cut-off of the KIVPA of 6 and higher

	Spearman correlation	Mean KIVPA for cases	Sensitivity	Specificity	PPV	NPV	AUC	95% CI
CBCL								
Total Problem score	0.41 *	7.08 * * * [a]	0.57 [c]	0.84 [e]	0.25	0.95	0.79	0.74 - 0.83
Total Problem score, not under treatment	0.40 *	6.94 * * * [a]	0.55 [c]	0.84 [e]	0.23	0.94	0.78	0.73 - 0.83
Internalizing	0.43 *	7.22 * *	0.63	0.84 [e]	0.26	0.96	0.81	0.77 - 0.86
Externalizing	0.30 *	5.68 * * * [a]	0.44 [d]	0.84 [e]	0.26	0.92	0.68	0.64 - 0.73
YSR								
Total Problem score	0.70 *	9.08 * * * [a]	0.82 [c]	0.85 [e]	0.28	0.99	0.92	0.90 - 0.95
Total Problem score, not under treatment	0.69 *	9.07 * * * [a]	0.82 [c]	0.86 [e]	0.29	0.97	0.93	0.90 - 0.95
Internalizing	0.72 *	9.36 * * * [b]	0.86 [d]	0.85 [e]	0.31	0.99	0.93	0.91 - 0.96
Externalizing	0.48 *	7.34 *	0.62	0.83 [e]	0.19	0.96	0.81	0.76 - 0.86
Referred because of psychosocial problems	-	7.04 * *	0.55	0.83 [e]	0.12	0.98	0.77	0.70 - 0.84

*   Statistically significant (p < 0.001) differing from null.
*   Overall mean 3.08; p-value of difference with other adolescents: < 0.001, t-test.
a: Differences between cases (adolescents with a clinical score) and others are larger for girls than for boys (p < 0.05; F-test in ANOVA).
b: Differences between cases (adolescents with a clinical score) and others are larger for girls than for boys (p < 0.01; F-test in ANOVA).
c: Sensitivity is higher for girls than for boys (p < 0.01; chi-square test).
d: Sensitivity is higher for girls than for boys (p < 0.001; chi-square test).
e: Specificity is lower for girls than for boys (p < 0.001; chi-square test).
PPV = positive predictive value; NPV = negative predictive value; AUC = area under the ROC curve; CI = confidence interval.

Table 4   *Added value of the Total KIVPA score in detecting mental health problems as measured by the CBCL and YSR, measured by the odds ratio (OR) for a clinical CBCL or YSR score in the case of an elevated KIVPA score. Crude OR and OR after adjustment for relevant socio-demographic characteristics (Adj.*), for all children and for children with and without CHP-identified mental problems; non-treated sample n=1,226*

	Crude OR	95% CI	Adj. OR*	95% CI
All children				
CBCL				
Total Problem score	6.59	4.29 - 10.13	6.87	4.37 - 10.81
Internalizing	8.32	5.31 - 13.05	9.94	6.13 - 16.14
Externalizing	3.78	2.58 - 5.54	4.18	2.79 - 6.28
YSR				
Total Problem score	26.70	14.89 - 47.85	32.38	17.09 - 61.36
Internalizing	35.61	18.88 - 67.18	57.96	28.01 - 120.0
Externalizing	8.28	5.07 - 13.53	8.72	5.20 - 14.63
Children without CHP identified problems				
CBCL				
Total Problem score	5.41	2.98 - 9.81	5.17	2.74 - 9.78
Internalizing	8.86	4.77 - 16.44	10.76	5.49 - 21.12
Externalizing	3.74	2.00 - 5.69	3.70	2.12 - 6.49
YSR				
Total Problem score	24.64	12.05 - 40.50	27.30	12.66 - 58.90
Internalizing	34.20	14.62 - 80.02	53.41	20.51 - 139.1
Externalizing	7.48	4.17 - 13.42	7.54	4.06 - 14.01
Children with CHP identified problems				
CBCL				
Total Problem score	4.07	2.06 - 8.03	5.04	2.41 - 10.56
Internalizing	3.71	1.87 - 7.35	4.77	2.24 - 10.17
Externalizing	2.60	1.22 - 4.19	2.64	1.37 - 5.12
YSR				
Total Problem score	18.18	6.20 -53.35	35.87	8.21 – 156.80
Internalizing	18.66	7.02 - 49.65	37.65 [a]	10.73– 32.10
Externalizing	12.02	3.47 – 41.57	19.58 [b]	4.26 – 90.10

OR = odds ratio; CI = confidence interval.

*: Adjusted for the following characteristics (between brackets categories and numbers of adolescents): gender (boys/587; girls/639), age (11-13 years/752; 14-16 years/474), at least one parent Dutch-born/1,156; other/70; family situation (two parents/1,103; one parent/108; other/15), number of siblings (one or more/1,136; none/87; unknown/3), parental educational level ((very) low/439; higher/635; unknown/152), parental employment status (at least one parent works > 16 hours per week/1,150; other/66; unknown/10); urbanization (not or mildly urbanized/969; (very) urbanised/251; unknown/6).

a: Higher added value in girls (p=0.004, change in deviance (-2log likelihood) between models with and without interaction of gender and KIVPA score, which follows a chi-square distribution).

b: Higher added value in girls (p=0.004, change in deviance (-2log likelihood) between models with and without interaction of gender and KIVPA score, which follows a chi-square distribution).

45

3.4 Discussion and conclusions

The Short Indicative Questionnaire for Psychosocial problems among Adolescents (KIVPA) has been developed to support the identification of psychosocial problems by CHPs. This study examined some of its psychometric qualities, and assessed whether it is suitable for and adds to the early detection of psychosocial problems among adolescents in CHP practice. Results show that the KIVPA measures one dominant construct but contains several redundant items. Despite this, the Total KIVPA score discriminates between adolescents with and without problems as measured by CBCL, YSR and treatment status. Using the YSR as criterion, sensitivity and specificity are good. However, the KIVPA is most sensitive for Internalizing Problems, and scores on it are much higher for girls. Finally, most adolescents easily fill out the KIVPA and it provides additional information on the occurrence of psychosocial problems among adolescents, again mostly regarding internalizing ones.

Methodology

Methodological factors are unlikely to have affected these results. In general, response was very high (89.8%) and representative for the Dutch population, and the same applied to the adolescents for whom all data were available (77.8%). Regarding validity, two well-validated questionnaires were used as criteria, the CBCL and the YSR, as well as information on referral by professionals. Because of complexity and high costs, structured clinical interviews such as the Diagnostic Interview Schedule for Children were not used as criterion.[25] This might have provided additional information, but seems to be rather similar to questionnaire-based information.[26] Finally, this study mostly concerned adolescents aged 13 and 14 (89%) who filled out the KIVPA at home. The original validation study partially concerned adolescents who filled it out in class (47%). This latter group reported more problems, according to the authors partially due to a much higher mean age (58% aged 15-18 years).[14] It is unlikely, however, that the inclusion of this latter group in the present study would yield different findings.

Usefulness for CHP practice

The usefulness of the KIVPA depends on three factors: its psychometric properties, its suitability in daily practice, and its added value regarding the detection of psychosocial problems among adolescents. Regarding psychometric properties, the KIVPA seems to be a one-dimensional construct but its main factor explains relatively little (18.9%) of the overall variance and 12 of its 31 items have low loadings (<0.40) on this factor, i.e. are redundant. Restriction to the other 19 items yields a similar reliability of the first component as for 31 items (0.83 vs. 0.84), which is acceptable,[27] and comparable with other questionnaires of similar size such as the Strengths and Difficulties Questionnaire.[28,29] Furthermore, the results of PCA, although very similar to those presented by the authors,[17] do not support the empirical existence of three separate dimensions within the concept measured by the KIVPA. The most likely explanation for this finding is that the questionnaire has not been developed to enable the discrimination of separate problem areas. Results do not support the weighting of items in the different subscales as advised by the authors.

Moreover, the contents of the questionnaire focus on internalizing problems. Correlation coefficients with both the YSR and CBCL are far better for this type of psychosocial problems than for externalizing problems; the latter may even be partially due to the fact that both types of problems sometimes occur together, diagnostic comorbidity.[30] The focus of the KIVPA on internalizing problems may also explain most of the differences observed between boys and girls. Additionally, KIVPA scores are in general associated more strongly with YSR scores than with CBCL scores. However, this can easily be explained by the well-known differences in information on the mental health of children that is provided by parents and children themselves.[31,32]

Differences between informants may also explain the more favourable results with the YSR than the CBCL regarding criterion validity. At the proposed cut-off of 6, validity and sensitivity are satisfactory with the YSR as criterion whereas they are rather low with the CBCL. At this cut-off, 99% of adolescents with a KIVPA score in the normal range also have a normal YSR score. However, only 28% of adolescents with an elevated Total KIVPA score have a clinical YSR Total Problem score (and even less regarding externalizing problems). Thus the majority of the adolescents with an elevated Total KIVPA score will be false positives. This shows that information on KIVPA scores can be used only in conjunction with other (clinical) information.

Regarding suitability, the KIVPA seems to be satisfactory. In this study, almost all (89.9%) participating adolescents filled it out completely. Moreover, previous studies show that CHPs can work with it quite well, and that filling it out and interpreting it takes little time (5-10 minutes and 2-3 minutes, respectively).[17] Finally, the information regarding the Total KIVPA score adds to the identification of psychosocial problems, especially internalizing ones, and interestingly it works best for adolescents in whom the CHP did not identify psychosocial problems.

In conclusion, the KIVPA appears to be suitable for CHP practice and to add to the identification of internalizing psychosocial problems. However, the questionnaire contains some redundant items and its cut-off needs further study, especially regarding differences by gender. The KIVPA may thus support the identification of internalizing psychosocial problems if its efficiency is improved. As such, the development of the instrument seems to have been guided mainly by the problems in CHP practice regarding the identification of psychosocial problems, and not *a priori* by a solid psychometric starting point (which has for instance guided the development of the CBCL). Crucial for the early detection of psychosocial problems is a solid clinical approach in daily CHP practice combined with a solid psychometric approach. Regarding the latter, an approach based on item response theory could lead to an improvement of the KIVPA in its present format.[33]

When using the KIVPA, the CHP needs additional sources of information for an appropriate assessment of externalizing psychosocial problems. In general, psychosocial problems cannot be identified solely on the basis of the KIVPA, without proper assessment of the adolescent by a professional, but neither can they on the basis of any other existing questionnaire.[34]

References

1   Costello EJ. Child psychiatric disorders and their correlates: a primary care pediatric sample. *J Am Acad Child Adolesc Psychiatry* 1989;28:851-5.

2   Zahner, GEP, Pawelkiewicz W, DeFrancesco JJ, Adnopoz J. Children's mental health service needs and utilization patterns in an urban community: An epidemiological assessment. *J Am Acad Child Adolesc Psychiatry* 1992;31:951-60.

3   Leaf PJ, Alegria M, Cohen P, et al. Mental health service use in the community and schools: results from the four-community MECA study. *J Am Acad Child Adolesc Psychiatry* 1996;35:889-97.

4   Verhulst FC, Van der Ende J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997;36:901-9.

5   Minister of Public Health. *Basistakenpakket jeugdgezondheidszorg [Listing of basic tasks for preventive child healthcare]*. The Hague: Ministry of Public Health and Sports, 2001.

6   Gregg J, Appleton RE. Community child health: an essential specialist service for the millennium. *Lancet* 1999;354(Suppl II):28-30.

7   Hirasing RA, Van Zaal MAE, Meulmeester JF, Verbrugge HP. *Child health in the Netherlands. Facts and figures*. Leiden: TNO Prevention and Health, 1997.

8   Burgmeijer RJF, Van Geenhuizen YM, Filedt-Kok Weimar T, De Jager AM. *Op weg naar volwassenheid: evaluatie jeugdgezondheidszorg 1996 [On the route to adulthood: evaluation preventive child healthcare 1996]*. Leiden: TNO Prevention and Health, 1997.

9   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adol Med* 2001;155:462-9.

10 Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Onderkenning van en beleid bij psychosociale problemen in de schoolgezondheidszorg [Identification of and management of psychosocial problems in school healthcare]. *Ned Tijdschr Geneeskd* 2001;145:1790-5.

11 Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: University of Vermont Department of Psychiatry, 1991.

12 Verhulst FC, Van der Ende J, Koot HM. *Handleiding voor de CBCL/4-18 [Manual for the CBCL/4-18]*. Rotterdam: Erasmus University/Department of Child and Adolescent Psychiatry, Sophia Children's Hospital, 1996.

13 Durlak JA, Wells AM. Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *Am J Community Psychol* 1998;26:775-802.

14 Bos CA, Van Ede J, Maarsingh EJ, Lako CL. De ontwikkeling van de KIVPA: de Korte Indicatieve Vragenlijst voor Psychosociale problematiek bij Adolescenten. [The development of the KIVPA: the Short Indicative Questionnaire for Psychosocial problems among Adolescents.] *Tijdschrift voor Jeugdgezondheidszorg* 1998;30:56-61.

15 Bouchier CAM, De Rover CM, De Vries-Lequin I, Kroesbergen HT, Visee WF, Aben DJM. Improving prevention in Dutch child healthcare in schools. *Eur J Public Health* 1999; 9: 200-204.

16 Maarsingh EJ, Van Ede J, Bos CA. Reactie van de werkgroep KIVPA op het artikel 'Is landelijk gebruik van de KIVPA verantwoord?' [Reaction of the KIVPA workgroup on the paper 'Is national use of the KIVPA questionnaire sound?'] *Tijdschrift voor Jeugdgezondheidszorg* 1999;31:67-8.

17 KIVPA working group. *Korte Indicatieve Vragenlijst Psychosociale problematiek Adolescenten: handleiding voor gebruik; analyses landelijk* bestand [Short Indicative Questionnaire for Psychosocial problems among Adolescents (KIVPA): manual for use; analyses on a national file]. Drachten: GGD, 1997.

18 Van der Linden FL, Dijkman TA *Jong zijn en volwassen worden in Nederland: een onderzoek naar het psychosociaal functioneren in alledaagse situaties van de Nederlandse jongeren in de leeftijd van 12 tot 21 jaar [To be young and grow up in the Netherlands: a study on psychosocial functioning in everyday situations of Dutch youth aged 12 to 21 years]*. Nijmegen: Hoogveld Institute, 1989.

19 Wiegersma PA, Van Kamp I. Is landelijk gebruik van de KIVPA verantwoord? [Is national use of the KIVPA questionnaire sound?] *Tijdschrift voor Jeugdgezondheidszorg* 1999;31:65-6.

20 COTAN (Commissie Testaangelegenheden Nederland [Committee on Testmatters Netherlands]). *Report on the KIVPA questionnaire.* Utrecht: COTAN, 2001.

21 Achenbach TM. Manual for the Youth Self Report and 1991 Profiles. Burlington: University of Vermont Department of Psychiatry, 1991.

22 Verhulst FC, Van der Ende J, Koot HM. *Handleiding voor de YSR [Manual for the YSR]*. Rotterdam: Erasmus University/Department of Child and Adolescent Psychiatry, Sophia Children's Hospital, 1997.

23 Rothman KJ, Greenland S. *Modern epidemiology*. Philadelphia: Lippincott, 1998.

24 SPSS for Windows, release 10.0.7. SPSS inc., 2000.

25 National Institute of Mental Health (NIMH). *Diagnostic Interview Schedule for Children (DISC)*. New York: New York State Psychiatric Institute, 1992.

26 Boyle MH, Offord DR, Racine YA, Szatmari P, Sanford M, Fleming JE. Adequacy of interviews vs checklists for classifying childhood psychiatric disorder based on parent reports. *Arch Gen Psychiatry* 1997;54:793-9.

27 Bland JM, Altman DG. Cronbach's alpha. *BMJ* 1997;314:572.

28 Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *Eur Child Adolesc Psychiatry* 1998;7:125-30.

29 Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 2000;177:534-9.

30 Angold A, Costello EJ, Erkanli A. Comorbidity. *J Child Psychol Psychiatry* 1999; 40: 47-87.

31 Verhulst FC, Ende van der J. Agreement between parents' reports and adolescents' self-reports of problem behavior. *J Child Psychol Psychiatry* 1992;33:1011-23.

32 Jensen PS, Rubio-Stipec M, Canino G, et al. Parent and child contributions to diagnosis of mental disorder: are both informants always necessary? *J Am Acad Child Adolesc Psychiatry* 1999;38:1569-79.

33 van Buuren S, Hopman-Rock M. Revision of the ICIDH Severity of Disabilities Scale by data linking and item response theory. *Stat Med* 2001;20:1061-76.

34 Hunter J, Higginson I, Garralda E. Systematic literature review: outcome measures for child and adolescent mental health services. *J Public Health Med* 1996;18:197-206.

4   Detecting psychosocial problems among 5–6-
    year-old children in Preventive Child
    Healthcare.
    The validity of a short questionnaire used in an
    assessment procedure for detecting
    psychosocial problems among children

T. Vogels

S.A. Reijneveld

E. Brugman

M. den Hollander-Gijsman

F.C. Verhulst

S.P. Verloove-Vanhorick

Abstract

Background:

An extended re-assessment of the psychometric properties of the LSPPK, an instrument aimed at identifying children with emotional and behavioural problems.

Methods:

Data came from a national sample in the Netherlands of parents of 1248 children (aged 5–6 years) interviewed by child health professionals (CHP). Data were obtained regarding psychosocial problems, treatment status and scores on the LSPPK (Parent and CHP Index), and on the Child Behavior Checklist (CBCL). The scale structure, reliability, criterion and content validity and added value of the LSPPK were assessed using the CBCL and treatment status as criteria.

Results:

The scale structure corresponded with that found originally; the LSPPK improved the prediction of problems according to the CBCL, compared to predictions using readily available risk indicators alone. Reliability varied between 0.55 and 0.69. For the LSPPK Parent Index, sensitivity varied between 0.42 and 0.50. For the CHP Index sensitivity varied between 0.60 and 0.96, but specificity varied between 0.76 and 0.79. Both indices were very sensitive for attention and social problems, but less so for other problems.

Conclusion:

The LSPPK Parent Index cannot distinguish sufficiently between children with or without serious problems. Either too many children with problems remain unnoticed or too many children without problems are labelled as a case. The LSPPK CHP Index, reflecting the CHP's interpretation of the Parent Index after interviewing the parents does not compensate adequately for the weaknesses of the Parent Index. Better assessment procedures and strategies need to be developed.

4.1 Introduction

A sizeable proportion of children suffers from behavioural and emotional problems without receiving treatment.[1–3] Such problems may have a negative impact on current functioning and future development. The persistence of such problems, until adulthood, has been well documented.[4,5] Durlak and Wells[6] showed that good screening procedures are essential for effective interventions. In the Netherlands, preventive Child Healthcare (CHC) is one of the most important low threshold services for the early detection of psychosocial. CHC is provided unasked to all children, from birth to 19 years, by community physicians and nurses (Child Health Professionals, CHP).[7,8] It provides publicly funded preventive programmes (screening, general physical examinations, vaccinations, health education and promotion). Over 90% of all children undergo three to four assessments by a CHP during their school careers. Nearly all services offer assessments to children in grade 2 of primary school (mean age: 5–6 years).[9] These assessments include a general physical examination, standardized screening procedures, and an interview with parents concerning health status and (physical, emotional and behavioural) developmental problems. The assessments take approximately 15 to 30 minutes. At the end of the assessment, the CHP decides whether counseling, follow-up, or referral are required. If CHPs identify serious problems, they have to refer the child to other professional services as they do not offer actual treatment themselves.

Despite a growing awareness of the importance of psychosocial problems for a healthy development, CHC's involvement in individual and collective interventions on psychosocial problems was hampered for a long time by the lack of good, reliable and valid assessment procedures. Such procedures must be manageable in the actual practice of preventive CHC in the limited time available for each individual child. Recently, several procedures were developed[10,11] claiming to offer a good, reliable and valid way of detecting children in need because of psychosocial problems. One such procedure is based on a short questionnaire, the LSPPK (acronym for 'Landelijk Signalerings-instrument Psychosociale Problematiek Kleuters', Dutch for 'National Checklist for Indicating Psychosocial Problems in Five/Six Year Olds').[11] The LSPKK is filled out by the parents before meeting the CHP (Parent Index, PI). Problems indicated by the parents are discussed with the CHP. After the meeting, the CHP indicates which problems are present in his/her view (Child Health Professional Index, CHPI) and which problems the parents still consider to be present. Many of the CHC institutes in the Netherlands adopted the LSPPK. An international publication on the LSPPK was published,[11] but as far as we know, the instrument is not yet used in other countries.

Bouchier et al[11] evaluated the LSPPK. They asked CHPs to judge the usefulness of the questionnaire and compared the CHPI and PI. Furthermore, they evaluated the psychometric properties of the PI. Correspondence between the conceptual scale structure and the (varimax rotated) factor matrix was shown to exist. Internal consistency was moderate (0.66 for the PI). The authors calculated a PI cut-off point, distinguishing optimally between children referred to mental health services and those not. The overall correct classification rate was 83%. The authors suggested that CHPs should seriously consider referring children with a higher score. The number of problems indicated by CHPs after the interview is much lower than that indicated by the

parents. Bouchier et al[11] interpret this difference as the result of the CHPs being able to reassure the parents.

Implicitly they assume that the professional's indication of problems is more accurate than parental assessment. Literature not always supports this view[12]; yet, it may be true. However, the authors did not present data to support this assumption. Their evaluation was also contaminated by the fact that their criterion (referral) is dependent on the information elicited using the LSPPK.

The aim of the current study was to replicate and to enhance the original psychometric evaluation in a large national sample. The LSPPK was validated on criteria that are independent from the LSPPK itself: is the child in question currently being treated by a mental health service, and a clinical score on the Child Behavior Checklist 4–18 (CBCL).[13] One of the primary objectives of CHC is to identify children with problems not yet being treated. Therefore, the relationship between CBCL and LSPPK was analysed for the group as a whole and for children not under treatment. Furthermore, as the original authors seem to consider the CHP's perception as more indicative of real problems than the perception of the parents, the current study extended the validation to include the CHPI.

4.2 Methods

4.2.1 Sample

The sample was obtained using a two-stage selection procedure. In the first stage, a random sample was taken, consisting of 19 of the 63 Child Healthcare Services, after stratification by region and degree of urbanization of their district. In the second stage, each service provided about 75 cases in the first grade of primary education by inviting all parents of children in three school classes to take part. Of all parents, 91% (n=1419) participated. Complete data for the crucial variables in this study (CBCL Total Problem score, PI, CHPI and being under treatment) were available for 1248 children (80%). Analyses were restricted to this group to make interpretation easier. Response was non-selective, details have been presented elsewhere.[1] Exclusion of children with incomplete data sets, however, resulted in the removal of relatively more children with serious problems. From the 63 children whose treatment status was not known, 22% had a higher than cut-off point score, compared to only 7% of those with known treatment status (p<0.001). From those children with an unknown CBCL Total score, 15% (n=3) had a higher than cut-off point score on the PI, compared to 8% of those with a known CBCL Total score (p<0.05). Known or unknowns scores on the PI or CHPI were not related to differences to CBCL and treatment status.

Details of the working sample are presented in Table 1.

Table 1   *Description of study group on demographic and family characteristics and problem indicators used in the study*

	Boys n=644 %	Girls n=604 %	Total n=1248 %
Gender	52	48	100
Age			
5/5 years	64	68	65
6/7 years	36	32	35
Member of ethnic minority	6	5	5
Number of parents in family			
2	92	91	91
1	6	7	6
Other	1	3	2
Only child			
Yes	8	10	9
No	91	90	91
Unknown	1	0	0
Highest education completed by parents			
1st phase secondary education or lower	33	30	32
Higher	67	69	68
Unknown	1	1	1
Employment status breadwinner			
Unemployed or <17 hours a week	6	7	7
Employed >16 hours a week	93	93	93
Unknown	1	0	0
Urbanization			
Living in (highly) urbanized area	70	69	69
Living elsewhere	29	30	30
Unknown	1	0	0
Being treated in mental health services			
No	92	95	94
Yes	2	2	2
Unknown	7	3	5
LSPPK Parent Index			
Mean 0–9 (sd)	1.3 (1.6)	1.1 (1.5)	1.2 (1.5)
% above cut-off point (>3)	9	6	8
LSPPK CHP Index			
Mean 0–9 (sd)	0.7 (1.3)	0.4 (1.0)	0.6 (1.2)
% above cut-off point (>0)	34	21	28
CBCL Total score			
Mean (sd)	17.1 (14.4)	15.1 (13.3)	16.1 (13.9)
% with a clinical score	8	8	8

NB Percentages do not always add up to 100% because of rounding.

4.2.2 Procedures and measures

The LSPPK and CBCL were sent to the parents along with the invitation to the regular check-up. At the time of the check-up, the CHP filled in a study-specific questionnaire, covering demographics and being under treatment. The LSPPK[11] is a nine-item instrument, covering three domains (behavioural, cognitive-developmental and emotional problems). Parents indicate whether problems described in the items (table 2) are present, selecting one of three categories ('not worried', 'worried a bit', 'very worried'). They submit the form to the CHP and problems indicated are discussed.

Following the check-up, the CHP indicates which problems are present in his or her view. A more detailed description of the instrument has been published by the original authors[11] This study re-evaluated the PI cut-off point calculated by Bouchier et al[11] For the CHPI no cut-off point was calculated. The authors informed us (written communication) that they felt that any problem indicated on the CHPI was an indication of a serious problem. This study therefore used a cut-off point of 0 vs. higher for the CHPI. The CBCL was used to assess the parent's report of the child's behavioural and emotional problems during the preceding six months.[14] The good reliability and validity of the CBCL established by Achenbach were confirmed for the Dutch translation.[13] The CBCL consists of 20 competence items and 120 problem items. Here only the problem items were used. Parents indicate the presence of problems, choosing one out of three categories (no problem, sometimes/a bit, often/a lot). Nine syndrome scales, two broadband groups of syndromes, designated Internalizing and Externalizing, and a Total Problem score were computed (see Table 5). Internalizing includes the Withdrawn, Somatic Complaints, and Anxious/Depressed syndrome scales, and Externalizing includes the Delinquent and Aggressive Behavior syndrome scales. Cases were subsequently allocated to a normal or a clinical range of the scoring distributions in the Dutch normative sample.[13] Cut-offs were set at the 97th percentile for the syndrome scales and at the 90th percentile for the Total Problem and broadband scales. CHPs participating in the study were not aware of the parents' answers on the CBCL.

The CHP registered whether or not a child was currently being treated by a mental health service, using both CHC files and information obtained during the checkup: 2% of all children were currently under treatment in various types of mental health services that offer counseling, psychological and/or psychiatric therapy, support for developmental problems and so on. We did not register the specific reasons for which they were being treated, but these vary from strictly psychological and psychiatric problems (e.g. lack of self-confidence or ADHD) to suffering from stressful circumstances (e.g. being bullied).

The criteria variables are, of course, empirically related. Of those currently under treatment, 46% has a clinical CBCL Total score. For those not under treatments this percentage is 7% (p<0.001).

4.2.3 Analysis

The psychometric properties of the LSPPK and its added value in identifying psychosocial problems were assessed. The procedures used by the original authors were first replicated.[11] To re-assess the correspondence between the conceptual domains and the scale structure, LSPPK's scale structure was assessed using principal component analysis (PCA, Eigenvalue >1, varimax rotation). Internal consistency was computed for the subscales, PI and CHPI. Criterion validity was then assessed using clinical CBCL Total Problem score and being under treatment for psychosocial problems as criteria and calculating ROC (Receiver Operating Characteristic) parameters (sensitivity, specificity and area under curve). These analyses were carried out for all children and for those not receiving treatment in mental health institutions. Thirdly, the content validity of the LSPPK Indices was assessed by calculating bivariate odds ratios of dichotomized LSPPK Indices with dichotomized CBCL syndrome and broadband scores. These

analyses were performed for all children and for those with and without a clinical CBCL Total Problem score, to determine which kind of problems occur relatively frequently among false positive and false negative results. The analysis of the total group as a whole provides information on the type of problems that the LSPPK is able to detect, indicated by strong associations. The analyses in the two subgroups would hopefully yield few statistical significant effects. Among the subgroup of children with a CBCL Total score in the normal range, the absence of such effects would indicate that the chance of getting classified was not related to specific problem domains. Similarly, for the group with a clinical CBCL Total score, a significant effect would indicate that the absence of a specific problem enhances the chance of a case not being detected.

Turning to the added value of the LSPPK, the extent to which the LSPPK improved the prediction of psychosocial problems, after using readily available risk indicators, was assessed. First, a logistic regression was performed using demographic and family characteristics. Subsequently, the dichotomized PI and CHPI were added to the model and the model's improvement was determined (change in –2 log likelihood and significance of change).

All analyses were conducted with SPSS 10.0. A significance level of 0.05 was used. Cases with missing values on the PI, CHPI and CBCL Total scales were excluded from the analysis. Cases with missing values on a predictor variable in the logistic regressions were included in the analysis by creating dummy variables.

The study was approved by the Medical Ethics Committee of TNO.

4.3 Results

4.3.1 Scale structure
Both for the PI and CHPI, the results of the PCA (table 2) showed a remarkable resemblance to the results reported by the original authors.[11] Most items had a loading of 0.40 or higher on the first principal component. Three principal components with an eigenvalue >1.0 were extracted and the loadings of the items on the varimax rotated factors resembled the conceptual scale structure.[11]

The PCA for the PI were based on the three-point items, as this was how Bouchier et al.[11] performed their analyses. Scale scores, however, are simple sum scores of dichotomized items. A replication of the analyses on dichotomous items showed a comparable factor structure, with slightly lower loadings. Cronbach's α were also somewhat lower (table 2).

Table 2   *LSPPK's factor structure; first principal component extracted (PC), varimax rotated factors and Cronbach's α for the LSPPK (sub) scales; Parent Index and CHP Index*

	Supposed subscale [a]	Parent Index				CHP Index			
		First PC	Varimax rotated factors			First PC	Varimax rotated factors		
			1	2	3		1	2	3
1 Does not play well with other children	B	0.59	0.55[b]	0.26	0.17	0.53	0.49[b]	0.08	0.36
2 Demands attention in annoying way	B	0.60	0.84[b]	0.02	0.10	0.65	0.84[b]	0.07	0.12
3 Stubborn	B	0.57	0.82[b]	0.05	0.03	0.65	0.85[b]	0.16	0.01
4 Problems with schoolwork	C	0.51	0.20	0.06	0.76[b]	0.52	0.25	0.08	0.71[b]
5 Speech problem	C	0.40	0.04	0.05	0.76[b]	0.38	-0.01	0.05	0.81[b]
6 Needs parents presence	E	0.32	-0.05	0.65[b]	-0.11	0.38	0.10	0.63[b]	-0.18
7 Shy/withdrawn timid behaviour	E	0.52	0.04	0.55[b]	0.36	0.40	-0.11	0.63[b]	0.19
8 Lack of self-confidence	E	0.63	0.15	0.70[b]	0.22	0.67	0.17	0.70[b]	0.30
9 Worries, fears	E	0.58	0.24	0.70[b]	-0.02	0.61	0.26	0.70[b]	0.02
		0.55				0.57			
Cronbach's á (3 categories)		0.67	0.64	0.58	0.45	0.69	0.67	0.49	0.59
Cronbach's á (2 categories)		0.63	0.57	0.54	0.36				

a: B=Behavioural; C=Language/Educational; E=Emotional
b: Loading on factor corresponding with presupposed sub-domain.

58

Table 3   *ROC co-ordinates for LSPPK (Parent Index and CHP Index) with criterion 'currently being treated in mental health institutions'; sensitivity (Sens), 1-specificity (1-Spec)and area under the ROC curve (AUC)*

	Parent Index				CHP Index			
Cut-off <	Sens	95% CI	1-Spec	95% CI	Sens	95% CI	1-Spec	95% CI
1	0.95	0.75–1.00	0.52	0.49–0.51	0.96	0.75–1.00	0.24	0.21–0.26
2	0.91	0.69–0.98	0.29	0.26–0.31	0.86	0.64–0.69	0.13	0.11–0.15
3	0.77	0.54–0.91	0.14	0.12–0.16	0.59	0.37–0.63	0.05	0.04–0.07
4	0.50	0.28–0.71	0.06	0.04–0.07	0.36	0.18–0.59	0.02	0.01–0.03
5	0.46	0.25–0.67	0.03	0.02–0.04	0.27	0.12–0.50	0.01	0.01–0.02
6	0.32	0.15–0.59	0.01	0.01–0.02	0.18	0.06–0.41	0.00	0.00–0.01
7	0.23	0.19–0.46	0.00	0.00–0.01	0.09	0.02–0.31	0.00	0.00–0.01
8	0.05	0.00–0.25	0.00	0.00–0.01	0.00	0.00–0.19	0.00	0.00–0.00
9	0.05	0.00–0.25	0.00	0.00–0.00	0.00	0.00–0.19	0.00	0.00–0.00
10	0.00	0.00–0.00	0.00	0.00–0.00	0.00	0.00–0.00	0.00	0.00–0.00
AUC	0.89	0.82–0.96			0.92	0.86–0.98		

Table 4   *ROC co-ordinates for LSPPK (Parent Index and CHP Index) with criterion 'clinical CBCL Total score'; sensitivity (Sens), 1-specificity (1-Spec), and area under the ROC curve (AUC); for all children and those currently not under treatment*

	Parent Index				CHP Index			
Cut-off <	Sens	95% CI	1-Spec	95% CI	Sens	95% CI	1-Spec	95% CI
All children								
1	0.95	0.88–0.98	0.51	0.46–0.52	0.69	0.59–0.78	0.24	0.22–0.27
2	0.87	0.78–0.93	0.27	0.25–0.30	0.59	0.49–0.69	0.13	0.11–0.15
3	0.72	0.62–0.80	0.13	0.11–0.15	0.43	0.33–0.53	0.05	0.04–0.06
4	0.49	0.39–0.59	0.04	0.03–0.06	0.22	0.15–0.32	0.02	0.01–0.03
5	0.32	0.23–0.42	0.02	0.01–0.03	0.16	0.10–0.25	0.01	0.00–0.01
6	0.19	0.12–0.28	0.01	0.00–0.01	0.09	0.04–0.17	0.00	0.00–0.01
7	0.11	0.06–0.19	0.00	0.00–0.01	0.05	0.02–0.12	0.00	0.00–0.00
8	0.04	0.01–0.11	0.00	0.00–0.00	0.01	0.00–0.06	0.00	0.00–0.00
9	0.01	0.00–0.06	0.00	0.00–0.00	0.00	0.00–0.00	0.00	0.00–0.00
10	0.00	0.00–0.00	0.00	0.00–0.00	0.00	0.00–0.00	0.00	0.00–0.00
AUC	0.87	0.83–0.91			0.77	0.71–0.83		
Children not under treatment								
1	0.94	0.85–0.98	0.49	0.46–0.52	0.60	0.48–0.71	0.21	0.19–0.24
2	0.84	0.74–0.91	0.25	0.22–0.27	0.49	0.38–0.61	0.10	0.08–0.12
3	0.66	0.54–0.76	0.11	0.08–0.13	0.35	0.25–0.49	0.03	0.02–0.04
4	0.42	0.31–0.53	0.03	0.02–0.04	0.14	0.08–0.25	0.01	0.01–0.02
5	0.22	0.14–0.33	0.01	0.01–0.02	0.09	0.04–0.18	0.00	0.00–0.01
6	0.10	0.05–0.20	0.00	0.00–0.01	0.04	0.01–0.12	0.00	0.00–0.01
7	0.05	0.02–0.13	0.00	0.00–0.01	0.03	0.00–0.10	0.00	0.00–0.00
8	0.01	0.00–0.08	0.00	0.00–0.00	0.00	0.00–0.05	0.00	0.00–0.00
9	0.00	0.00–0.05	0.00	0.00–0.00	0.00	0.00–0.05	0.00	0.00–0.00
10	0.00	0.00–0.00	0.00	0.00–0.00	0.00	0.00–0.00	0.00	0.00–0.00
AUC	0.86	0.82–0.91			0.73	0.67–0.89		

4.3.2 Criterion validity

Table 3 presents the ROC parameters for the PI and CHPI, in relation to being under treatment for psychosocial problems or not. The area under curve (AUC) for the PI was 0.89; sensitivity at the cut-off point calculated by the original authors was 0.50; specificity was 0.94. So, half of those under treatment scored under the cut-off point, but only 6% of those not under treatment had an elevated score. For the CHPI, the AUC was 0.92, sensitivity at a cut-off point of >0 is high (0.96); specificity was 0.76.

For the second criterion – clinical CBCL Total problem score – the PI parameters were comparable (Table 4). For the CHPI, AUC and sensitivity were lower, with 69% of those with a clinical CBCL Total problem score detected at cut-off point. For children currently not under treatment, sensitivity of the PI and CHPI was somewhat lower.

4.3.3 Content validity

Table 5 presents the results of the evaluation of the content validity of the LSPPK using a clinical CBCL Total Problem score as a criterion. For the group as a whole clinical scores on all syndrome scales were related to a significantly higher chance of a PI score above cut-off point. By far the strongest relationship was found with clinical scores for Social Problems and Attention Problems. Clinical scores for Internalizing Problems were more closely related to a high PI score than Externalizing Problems.

Associations between the dichotomized CHPI and clinical CBCL scores are somewhat weaker. Here, the strongest relationship was found for Social Problems and for Anxious/Depressed. It should be noted that all children with a clinical score for Attention Problems had an elevated CHPI score. Therefore, odds ratio (OR) could not be calculated. Internalizing Problems were more strongly related to a CHPI score above the cut-off point than Externalizing Problems.

The analyses were replicated for children with a nonclinical CBCL Total Problem score (table 5, part B). This analysis allowed for a comparison between true negative and false positives, using a clinical CBCL Total Problem score as the criterion. For the PI a clinical score for Attention Problems is very strongly related to a higher risk of belonging to the false positives. Internalizing Problems also enhance this risk, though to a far smaller extent. Clinical scores on Thought Problems and to a lesser extent Internalizing and Externalizing are related to an increased risk of getting an elevated CHPI score, despite a CBCL Total Problem score in the normal range. All children with a clinical score on this scale had an elevated CHPI-score, despite a non-clinical CBCL Total Problem score.

Part C of the Table compares false negatives to true positives, again using the CBCL Total Problem score as the criterion. The likelihood of an elevated PI score was related to clinical Social Problems and Attention Problem scores. In other words, a non-clinical score on these scales is significantly related to a lower likelihood of an elevated PI score, despite a clinical CBCL Total Problem score. For the CHPI no significant relations were found. Once again, for the Attention Problems scale, an OR could not be calculated.

4.3.4 Added value of the LSPPK

In bivariate logistic regression analyses of all demographic and family characteristics presented in Table 1, only family status and employment status showed a significant relationship to any of the psychosocial criteria. Adding PI in the analysis resulted in an OR of 20.65 (95% confidence interval (CI): 11.58–36.82) for this factor. Removing it would result in a significantly worse prediction (p<0.001). For CHPI the OR is lower (5.33 (CI: 3.29–8.64) but removal would again significantly worsen the model's predictive power (p<0.001).

Table 5   *Results of bivariate logistic regression of clinical scores on CBCL syndrome and broadband scales on LSPPK Parent Index (<4 versus >3) and CHP Index (0 versus >0); OR and 95% confidence intervals (CI); overall and by CBCL Total Problem score (non-clinical versus clinical)[a]*

	Part A All children		Part B Non-clinical CBCL Total Problem Score		Part C Clinical CBCL Total Problem Score	
	OR	95% CI	OR	95% CI	OR	95% CI
LSPPK PI						
Clinical CBCL syndrome score:						
Withdrawn	14.43	5.44–38.32				
Somatic complaints	6.24	2.46–15.85				
Anxious/depressed	21.66	7.69–60.99				
Social problems	240.87	31.68–1831.40			22.05	2.78–74.78
Thought problems	9.57	4.26–21.46				
Attention problems	60.11	19.77–182.81	71.61	7.31–701.71	6.40	1.71–23.98
Delinquent behavior	4.70	1.64–13.47				
Aggressive behavior	12.61	5.89–27.00				
Sexual problems	3.46	1.46–8.22				
Internalizing	15.18	9.23–24.95	7.80	3.17–9.19		
Externalizing	8.70	5.39–14.06				
n	1248		1148		100	
n (PI >3)	98		49		49	
LSPPK CHPI						
Clinical CBCL syndrome score:						
Withdrawn	6.42	2.25–18.36				
Somatic complaints	5.37	2.15–13.42				
Anxious/depressed	18.87	4.27–83.42				
Social problems	21.67	4.96–94.61				
Thought problems	4.31	1.94–9.59	7.95	1.53–41.20		
Attention problems	[b]		[b]		[b]	
Delinquent behaviour	3.31	1.30–8.46				
Aggressive behaviour	6.06	2.73–13.45				
Sexual problems	2.35	1.16–4.76				
Internalizing	6.04	3.79–9.61	4.29	2.06–8.96		
Externalizing	4.07	2.70–6.14	2.04	1.03–4.05		
n	1248		1148		100	
n(CHPI>3)	347		278		69	

a: Results are only reported when 95% CI does not include 1.
b: OR cannot be calculated because all children with a clinical Attention Problems score have a CHPI-score >0.

4.4 Discussion

This study reassessed the psychometric properties of the LSPPK, a questionnaire-based procedure for detecting psychosocial problems among children aged 5 and 6. The factor structure clearly reflected the supposed conceptual structure. Adding the PI and CHPI to a predictive model, together with readily available demographic predictors, clearly improved the prediction. Cronbach's α varied between 0.55 and 0.69. The sensitivity of the PI at the cut-off point calculated by the original authors varied between 0.42 and 0.50. The sensitivity of the CHPI at the cut-off point suggested by the authors varied

strongly, depending on the criterion variable used (0.60–0.96). Specificity varied between 0.76 and 0.79. Clinical scores on most CBCL syndrome scales are associated with a higher chance of elevated PI and CHPI scores, but Attention Problems and Social Problems seem to be weighted much more heavily than other problems.

The data on which this re-evaluation was based were collected in a representative sample from the target population of the LSPPK. Data collection, furthermore, took place in the same setting for which the LSPPK was developed. The results of these analyses, therefore, may be expected to be valid for the field in which the LSPPK is being used, the Dutch Preventive Child Healthcare system.

The evaluation of questionnaires for emotional and behavioural problems is always hampered by the lack of a golden standard: there is simply no definitive indicator of such problems. This study therefore adopted a common strategy to overcome this problem: the use of different criterion variables and validation of a short questionnaire on a longer and widely accepted questionnaire. The following criteria were used: a (widely accepted and validated) clinical CBCL Total Problem score, receiving treatment from a mental health service and a clinical CBCL Total score whilst not under treatment. We could not use a psychiatric interview in this study, because of costs and burden for the parents. Such an interview might have added, though it is not always better than questionnaire-based information.[15]

The results of the analyses using these criteria were in general quite comparable, which suggests that the conclusions are robust. The high sensitivity of the CHPI for the criterion 'under treatment' is an exception, though. This high sensitivity is most likely to be explained by the low cut-off point used in the analysis, which leads to a quarter of the sample being labelled as a suspected case. This cut-off would also lead to 24% of the non-cases being labelled as suspected cases, resulting in enormous costs, both financial and psychological.

The analysis showed that at a group level the PI and the CHPI have added value in terms of distinguishing between children with and without problems. However, Cronbach's α values of the PI and CHPI were well below 0.90, the minimal value deemed necessary to justify scales as a selection instrument, e.g. for deciding whether individual children are in need of further attention or not.[16]

Due to the low cut-off point the sensitivity of the CHPI is higher than that of the PI, however, at the cost of specificity. This sheds doubt on the original authors' implicit assumption that the CHPI should be considered as the better indicator. The ROC parameters presented showed that simply increasing or decreasing the cut-off point is not a solution, Increasing the specificity by means of a higher cut-off point would lead to what Glascoe and Dworkin call a significant under-identification of children with behavioural and emotional problems.[12]

The high number of cases 'detected' by the LSPPK without a clinical CBCL Total score and without being under treatment might be due to the fact that the LSPPK covers a partially different domain of problems. This especially holds for its two items that relate to the Language/Education domain. To examine this possible explanation, we analysed whether these items had a high association with an elevated PI and CHPI score, among those with a normal range CBCL Total score. Results showed that this was not the case:

among the three items with the strongest association in this subgroup, only one came from the Language/Education domain. Similar results were obtained within the group currently not being treated. The language / education domain in the LSPPK is therefore not to be held responsible for the high number of false positives.

The content evaluation showed that some problem areas, especially Social Problems and Attention Problems, were much more likely to be associated with elevated LSPPK scores than others. The same syndrome scales were associated with a higher likelihood of a false positive result and of a false negative result. This suggests that such problems may be reflected too heavily in the LSPPK. Similar results were obtained in analyses, not shown in this article, which used different cut-off points for the LSPPK-Indices.

Preventive Child Healthcare needs a sensitive, valid, reliable and manageable assessment procedure for emotional and behavioural problems. Relying on the LSPPK Indices alone will lead to the detection of many children with Attention and Social Problems. Children with other problems, however, may very well remain unnoticed. In so far as the CHPI may be seen as the result of the interview between parents and CHP, the results presented here suggest that the interview, as it is conducted now, does not offer a sufficient solution.

It may be argued that the LSPPK does not perform significantly worse than other short questionnaires, such as the Strength and Difficulties Questionnaire.[17] This may be true, but it does not mean that the weakness of the procedure is acceptable. Glascoe[18] shows that children with false-positive scores on several screening tests perform significantly worse than true negatives and that these children therefore need special attention. This may be true, but his argument does not take into account the possible negative effects and the costs involved. A large number of false positives in a population-based screening program will lead to huge financial costs. Furthermore, it may have negative impacts for all those involved; many parents will be unnecessarily worried; children might be stigmatized and health services referring many children without reason to other (mental) health services, will hamper good working relationships with those services. Finally, in the Netherlands as in many other countries, only those screening programmes are deemed permissible that meet the strictest criteria of sensitivity and specificity.

Therefore, in our view improvements in the quality of the detection are necessary. The content evaluation in this article suggests that an adaptation of the content of the items is advisable.

The literature suggests that using more sources (e.g. teachers, the child itself), wherever possible, really improves the accuracy of screening and detection procedures.[17] However, such a strategy is very time consuming and therefore less viable in situations where large groups of children are to be assessed. In our view two alternative strategies might be far more promising. The first strategy is a detection procedure in more phases. In the first phase, short and highly sensitive instruments should be used. In a second phase, suspected cases should be assessed more intensively. A second strategy, possibly in combination with the first one, is offered by new testing techniques, such as computerized adaptive testing and Item Response Theory.[18] Such testing techniques allow for sensitive and specific testing with just a few items being asked in each individual case. Essentially, this is done by choosing items to be offered based on the

preceding answers. As far as we know, such techniques are absent in the field of emotional and behavioural problems and developing them would require investment. The cost of such an investment, however, will be far less then the financial and human costs of both under- and overdetection of social and behavioural problems.

References

1   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001;155:462-9.

2   Costello EJ. Child psychiatric disorders and their correlates: a primary care pediatric sample. *J Am Acad Child Adolesc Psychiatry* 1989;28:851-5.

3   Verhulst FC, Van der Ende J. Factor associated with child mental health service use in the community. J Am Acad Child Adolesc Psychiatry 1997;36:901-9.

4   Verhulst FC, Van der Ende J. The eight year stability of problem behavior in an epidemiologic sample. *Pediatr Res* 1995;38:612-7.

5   Ferdinand RF, Verhulst FC, Wiznitzer M. Continuity and change of self-reported problem behaviors from adolescence into young adulthood. *J Am Acad Child Adolesc Psychiatry* 1995;34:680-90.

6   Durlak JA, Wells AM. Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *Am J Community Psychol* 1998;26:775-802.

7   Hirasing RA, Van Zaal MAE, Meulmeester JF, Verbrugge HP. *Child health in the Netherlands: facts and figures*. Leiden: TNO Prevention and Health, 1997.

8   Verloove-Vanhorick SP. *Maternal and child care in different public health programmes*. XXII International Congress of Pediatrics. Amsterdam, 9-14 August 1998:47.

9   Burgmeijer RJF, Van Geenhuizen YM, Filedt-Kok Weimar T, De Jager AM. *Op weg naar volwassenheid: evaluatie jeugdgezondheidszorg 1996 [On the route to adulthood: evaluation preventive child Healthcare 1996]*. Leiden/Maarssen: TNO Prevention and Health / KPMG Consultants, 1997.

10   Maarsingh EJ, Van Ede J, Bos CA. Reactie van de werkgroep KIVPA op het artikel 'Is landelijk gebruik van de KIVPA verantwoord?' [Reaction of the KIVPA working group on the article 'Is nationwide use of the KIVPA questionnaire sound?']. *Tijdschr Jeugdgezondheidszorg* 1999;31:67-8.

11   Bouchier CAM, De Rover CM, De Vries-Lequin I, Kroesbergen HT, Visee HT, Aben DJM. Improving prevention in Dutch healthcare in schools. *Eur J Public Health* 1999;9:200-4.

12   Glascoe FP, Dworking PH. The role of parents in the detection of developmental and behavioral problems. *Pediatrics* 1995;6:829-36.

13   Verhulst FC, Van der Ende J, Koot HM. *Handleiding voor de CBCL/4-18 [Manual for the CBCL/4-18]*. Rotterdam, the Netherlands: Erasmus Universiteit/ Dept of Child and Adolescent Psychiatry, Sophia Children's Hospital, 1996.

14 Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile.* Burlington: University of Vermont Department of Psychiatry, 1991.

15 Boyle MH, Offord DR, Racine YA, Szatmari P, Sanford M, Fleming JE. Adequacy of interviews vs checklists for classifying childhood psychiatric disorder based on parent reports. *Arch Gen Psychiatry* 1997;54:793-9.

16 Bland JM, Altman DG. Statistics Notes: Cronbach's α. *BMJ* 1997;314:572.

17 Goodman R, Ford T, Simmons H, Gatward R, Metltzer H. Using the Strength and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 2000;177:534-9.

18 Glascoe FP. Are overreferrals on developmental screening tests really a problem? *Arch Pediatr Adolesc Med* 2001;155(1):54-9.

19 Wainer H, editor. *Computerized adapitive testing, a primer*. Hillsdale, 1990.

5 Use of the Pediatric Symptom Checklist for the detection of psychosocial problems in Preventive Child Healthcare

S.A. Reijneveld

A.G.C. Vogels

F. Hoekstra

M.R.Crone

Abstract

Background:

Early detection and treatment of psychosocial problems by preventive child healthcare may lead to considerable health benefits, and a short questionnaire could support this aim. The aim of this study was to assess whether the Dutch version of the US Pediatric Symptom checklist (PSC) is valid and suitable for the early detection of psychosocial problems among children.

Methods:

We included 687 children (response 84.3%) aged 7–12 undergoing routine health assessments in nine Preventive Child Health Services across the Netherlands. Child health professionals interviewed and examined children and parents. Before the interview, parents completed an authorised Dutch translation of the PSC and the Child Behavior Checklist (CBCL). The CBCL and data on the child's current treatment status were used as criteria for the validity of the PSC.

Results:

The consistency of the Dutch PSC was good (Cronbach alpha 0.89). The area under the ROC curve using the CBCL as a criterion was 0.94 (95% confidence interval 0.92 to 0.96). At the US cut-off (28 and above), the prevalence rate of an increased score and sensitivity were lower than in the USA. At a lower cut-off (22 and above), sensitivity and specificity were similar to that of the US version (71.7% and 93.0% respectively). Information on the PSC also helped in the identification of children with elevated CBCL Total Problems Scores, above solely clinical judgment.

Conclusion:

The PSC is also useful for the early detection of psychosocial problems in preventive child healthcare outside the USA, especially with an adjusted cut-off.

5.1 Background

Early detection and treatment of psychosocial problems may lead to considerable health benefits. Psychosocial problems have a high prevalence rate and lead to high costs of disease.[1] They also cause substantial restrictions in daily functioning in later life and are the major cause of long-term work disability in young adults.[1] Only a minority of children with psychological or psychosocial problems are under treatment.[2-4] If untreated, problems are likely to persist in later life and can lead to serious limitations in daily functioning.[2,5] Research has shown that early detection and treatment improves these children's prognosis substantially,[6,7] but a complete analysis of its cost effectiveness has yet to be carried out.

The community child health service is an ideal setting for the early detection of psychosocial problems among children as routine health examinations are provided through it for the entire population, as a standardised part of preventive child healthcare (PCH). In the Netherlands, municipalities are obliged by law to guarantee proper access to this type of care, free of charge.

However, the predictive value of early detection of psychosocial problems by PCH is still too low.[3,4] For instance, Brugman et al. show that even though Dutch PCH identifies psychosocial problems in 25% of all children of school age, they miss 43% of the children with a clinical score on the Child Behavior Checklist (CBCL).[3] Similarly, Murphy et al. reported that paediatricians had identified psychosocial problems in less than half of the children with elevated scores on the Pediatric Symptom Checklist (PSC) or the Child Global Assessment Scale.[8]

The PSC is a 35-item parent-completed questionnaire that supports the identification of psychosocial problems by paediatricians.[8-14] It takes less than 5 minutes to complete and score, and reflects the parent's impression of his or her child's psychosocial functioning. Its validity has been demonstrated in various paediatric settings in the USA, nationally,[10] in inner-city children,[8] in Hispanic children[11,12] and in children of substance-abusing parents.[14] Moreover, the PSC has recently been used as an outcome measure in the assessment of interventions to reduce the impact of trauma.[13] Given its good validity and applicability in US community child health services, the PSC is a likely candidate for use in other countries with similar systems of preventive child healthcare, such as the Netherlands.

The aim of this study was to assess the test properties of the Dutch version of the PSC and determine whether it would be suitable for and contribute to the early detection of psychosocial problems in children aged 7–12 by PCH.

5.2 Methods

This study is based on a community sample of children for whom PSC and CBCL data are available, and data on the identification and management of psychosocial problems by CHPs.

5.2.1 Population

The sample was obtained using a two-stage selection procedure. In the first stage, a national sample from 9 of the 41 Dutch Preventive Child Health Services was taken. In the second stage, each Service provided a sample of children aged 7–12 who were invited for routine well-child examinations. We aimed at a sample size of 700 respondents for evaluation, as earlier studies[15,16] demonstrated that short questionnaires used in PCH settings allow for an area under the ROC curves (AUC) of about 0.90 with a clinical CBCL Total Problems score as criterion. A sample of 700 suffices to estimate this AUC with a 95% confidence interval of +/- 0.02.

Of the total sample of 815 eligible children, 687 participated and 674 provided complete data on both questionnaires (84.3% and 82.7% of the original sample, respectively). Both groups were representative of the total sample regarding age and gender, but non-response was higher for children of immigrant/minority origin (27.4% vs. 12.2%). Analyses were restricted to children with complete data for both questionnaires to make interpretation easier.

5.2.2 Data collection

The data were collected according to a standardised procedure during routine well-child examinations, from September 2004 to July 2005. The study was performed in compliance with the Helsinki Declaration.[17] The design of the study was approved by the local TNO Medical-Ethical Committee and includes verbal informed consent by parents.

The PSC[10] and the CBCL[18,19] were mailed to children, along with the standard invitation for the preventive health assessment. Before attending the assessments, parents completed the questionnaires, placed them in sealed envelopes and gave them to the CHPs, who in turn passed them on to the researchers without opening them (in contrast with routine use, where the CHP would partially base the interview on the information from the PSC). The CHP interviewed each child and its parents regarding mental health and background, and examined each child. After each assessment, the CHP answered the following question: 'Does the child have a psychosocial problem, at this moment?' (yes, no) and scored its severity (mild, moderate or severe) and the type of problems identified using a pre-coded list. Children who only had risk indicators for the development of psychosocial problems, such as having parents with psychiatric problems or other family problems, had to be coded as having no problems.

The PSC was translated following the procedure proposed by Guillemin et al.[20] Firstly, the original US English version of the questionnaire was translated into Dutch by three certified translators working independently of each other. Secondly, three further certified translators each translated one Dutch translation back into US English. The resulting US English versions were compared to the originals and all discrepancies were discussed by three researchers (SAR, MRC and AGCV) who spoke both Dutch and English. Discrepancies were also discussed with the developers of the PSC, Dr J.M. Murphy and Dr M.S. Jellinek, especially where items raised questions as to their intended meaning. The PSC consists of 35 items that are rated as never, sometimes or often present (0, 1 and 2, respectively). Item scores are summed; we dichotomised at 0–27 vs. 28–70, following the US cut-off.[10]

The CBCL was used to assess parents' reports of the behavioural and emotional problems of their children over the preceding six months. Its (good) reliability and validity has been established.[18,19] We used only the 120 problem items from the CBCL and computed scores for two broad-band groups of syndromes designated as Internalising and Externalising, and a Total Problems score. Children were also allocated to a normal range or a clinical range, using the 90th percentile of the Dutch normative sample as the cut-off. [19]

5.2.3 Analysis

In the analysis we assessed the psychometric properties of the PSC and its added value in identifying psychosocial problems. Regarding psychometric properties, we first computed its internal consistency and examined the fit between the scale structure and the observed data using confirmatory factor analysis (CFA) with structural equation modeling. Next, we assessed the validity of the PSC using dichotomised CBCL scores (Total Problems score and Internalising/Externalising scales) and referral by the CHP due to psychosocial problems as criteria. Finally, we assessed whether mean PSC scores differed with the children's background.

Regarding the added value of the PSC in identifying psychosocial problems, we assessed the odds of identification of mental health problems (i.e. a clinical CBCL Total Problems score) using an elevated score on the PSC. This was repeated with adjustment for social and demographic risk indicators known to the CHP that might have helped in the identification of psychosocial problems.[3,4] Regarding social and demographic risk indicators, we retained children with missing data in the logistic regression models by creating separate dummies for the missing category of each variable.

All analyses were done with SPSS 12.0 for Windows,[21] except the CFA, which was done with Amos 5.[22] All analyses were repeated for boys and girls separately. Results for these subgroups are provided only if they differed in a statistically significant way ($p < 0.05$).

5.3 Results

5.3.1 Demographics

The average age of the children in the study was 9.7 years (standard deviation 1.4 years) and there were slightly more girls than boys. Further demographic information is presented in Table 1.

5.3.2 Scores on PSC and CBCL

Mean scores on the PSC are slightly higher for boys than for girls, which also holds for the CBCL (Table 2a). The internal consistency of the PSC was very good (Cronbach's alpha 0.89), though the CFA revealed that the items could not be fully represented by a single factor (Chi-square = 2715 at 560 df; $p < 0.001$; Goodness-of-Fit Index (GFI) = 0.75; Parsimony corrected GFI = 0.66).

Table 2b shows the prevalence rates of elevated scores on the same questionnaires using their established cut-offs.[10,19] Of all the children, 4.5% had elevated scores on the PSC and 8.9% had elevated scores on the CBCL. The latter closely resembles its distribution in the Dutch normative sample. In US populations, the prevalence of elevated PSC scores ranges from 12–14%. This corresponds to a cut-off of 0–21 vs. 22+ among Dutch children, when compared in Table 2b. To enable comparisons with US data on the PSC, all further analyses are presented for this cut-off too.

Table 1:   *Demographic characteristics of the participating children, and mean Pediatric Symptom Checklist scores for selected sociodemographic groups (n = 674)*.*

Characteristic	No. (%) *	Mean	SD	P-value **
Gender				0.005
Male	327 (48.2%)	12.5	8.4	0.005
Female	347 (51.8%)	10.8	7.7	
Ethnic background				
Dutch	553 (91.3%)	11.3	7.8	0.034
Immigrant/minority	53 (8.7%)	13.7	8.5	
Family composition				<0.0001
Two parents	558 (85.5%)	10.9	7.7	
Single parent	78 (11.9%)	15.7	8.4	
Other	17 (2.6%)	13.8	7.1	
Highest parental education				0.57
Only primary school (=8 years)	21 (3.3%)	11.9	5.7	
Lower vocational (=max. 12 years)	182 (28.6%)	11.3	7.7	
Higher vocational (=max. 16 years)	196 (30.8%)	11.0	7.3	
University/higher professional (17 years and over)	238 (37.4%)	12.0	8.8	
Parental employment status				0.016
No paid employment	38 (5.6%)	15.1	11.7	
One parent with paid employment	193 (28.6%)	11.8	7.8	
Two parents with paid employment	443 (65.7%)	11.2	7.7	

* Numbers do not always total 674 because of missing values.
** P-value for differences in mean scores by background characteristic.

Table 2a   *Scores on the PSC and CBCL Total Problems, Internalising and Externalising scales, for all children and by gender (mean, standard deviation, range).*

Scale	Total n=674 mean (SD)	range	Boys n=327 mean (SD)	Range	Girls n=347 mean (SD)	range
PSC	11.6 (8.1)	0–46	12.5 (8.4)	0–40	10.8 (7.7)	0–46
CBCL						
Total	18.7 (15.3)	1–118	20.1 (16.7)	1–118	17.3 (13.9)	1–74
Internalising	5.4 (5.5)	0–35	5.3 (5.5)	0–35	5.6 (5.5)	0–31
Externalising	5.7 (6.0)	0–34	6.7 (6.8)	0–34	4.8 (5.0)	0–25

Table 2b   *Numbers and percentages of children with elevated scores on the PSC and CBCL*
Total Problems, Internalising and Externalising scales, and of children currently
*under treatment for psychosocial problems, for all children and by gender.*

Scale	Total (n = 674)	Boys (n = 327)	Girls n = 347
PSC[a]	30 (4.5%)	18 (5.5%)	12 (3.5%)
PSC adjusted[b]	86 (12.8%)	53 (16.2%)	33 (9.5%)
CBCL Total[c]	60 (8.9%)	32 (9.8%)	28 (8.1%)
Internalising[d]	82 (12.2%)	40 (12.2%)	42 (12.1%)
Externalising[e]	52 (7.7%)	28 (8.9%)	23 (6.6%)
Currently under treatment for psychosocial problems	53 (7.9%)	34 (10.4%)	19 (5.5%)

[a] 28 and over for boys and girls (i.e. original US cut-off)
[b] 22 and over for boys and girls (i.e. adjusted Dutch cut-off)
[c] 38 and over for boys and girls
[d] 13 and over for boys, and 16 and over for girls
[e] 14 and over for boys, and 13 and over for girls
All cut-offs refer to the age groups studied

5.3.3 Validity

Subsequently, the degree to which the score on the PSC is truly elevated in the case of psychosocial problems as measured by these four criteria (i.e. sensitivity) and the degree to which it is 'normal' in the case of the absence of these problems (i.e. specificity) were assessed. For the recommended cut-off of the PSC at 28 and above, scores were 0.33 and 0.98 respectively, using a clinical CBCL Total Problems score as the criterion, and 0.19 and 0.97 respectively, using being under treatment for mental health problems as the criterion. Figure 1 shows the Receiver Operating Characteristic (ROC) curve for all possible cut-off points. The curve is close to the upper-left corner of the figure, particularly when the CBCL is used as the criterion, indicating a high validity of the PSC if this gold standard is used. Curves for CBCL Internalising and Externalising Problems are largely similar but slightly more off the upper-left corner (i.e. less favourable; not shown). The same holds for problems detected by the CHP when compared with the curve for 'under treatment' (not shown). Table 3 shows the resulting areas under the ROC curves (AUC) and positive and negative predictive values for both cut-offs. Results regarding AUCs did not differ by gender (not shown).

Figure 1
Figure 1 Receiver Operating Characteristic (ROC) curve for all possible cut-off points of the Dutch version of the PSC, using a clinical CBCL score and Currently Under Treatment or Psychosocial Problems as criteria.
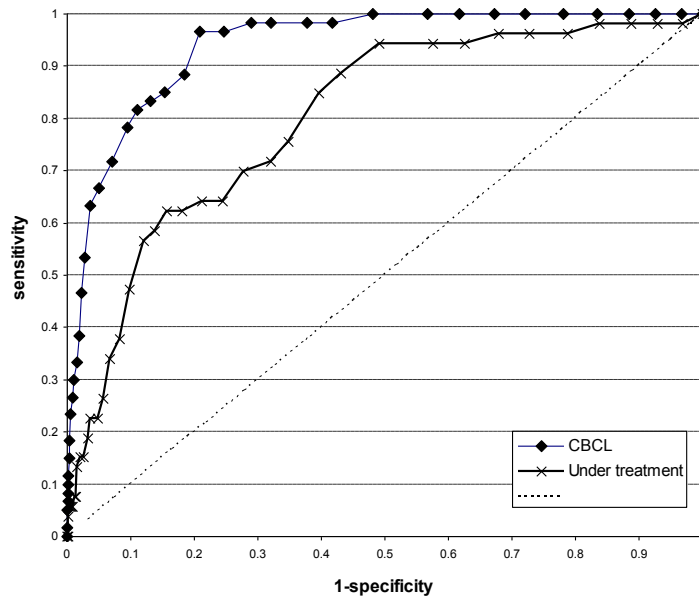
5.3.4 Differences in scores by background characteristics
Mean PSC scores were higher for boys, for children from minority backgrounds, single-parent families and unemployed families (Table 1, final columns).

5.3.5 Added value
Finally, we examined the degree to which information from the PSC contributed to the diagnosis of psychosocial problems as measured by the CBCL over and above the clinical opinion of the child health physician without knowledge of the PSC. This yielded an odds ratio of 21.3 (95% confidence interval 8.7 to 52.2), with the only predictive background characteristic being family composition. Using the alternative PSC cut-off of 22+ yielded slightly higher odds ratios.

5.3.6 Parent opinion of the PSC
A large majority of parents completed the PSC fully (91.1%) and no parent missed more than 3 items. However, 20% of parents made critical remarks about the PSC, mainly concerning lack of fit between questions and answer categories (7%) and unclear questions (5%).

Table 3   Sensitivity (Sens.), specificity (spec.), AUC and positive and negative predictive value (PPV, NPV) for the PSC at cut-off 27, using CBCL Total, Internalising and Externalising Problems, and under treatment for psychosocial problems as criteria (n = 674).

| | Sens. | Cut-off 0–27 vs. 28+ | | | Cut-off 0–21 vs. 22 | | | | AUC | 95% CI |
		Spec.	PPV	NPV	Sens.	Spec.	PPV	NPV		
CBCL										
Total	33.3%	98.4%	66.7%	93.8%	71.7%	93.0%	50.0%	97.1%	0.94	0.92–0.96
Internalising	24.4%	98.3%	66.7%	90.4%	53.7%	92.9%	51.2%	93.5%	0.89	0.86–0.92
Externalising	30.8%	97.7%	53.3%	94.4%	57.7%	91.0%	34.9%	96.3%	0.90	0.87–0.93
Currently under treatment	18.9%	6.8%	33.3%	93.3%	47.2%	90.2%	29.1%	95.2%	0.80	0.74–0.86

5.4 Discussion and conclusion

This study assessed the psychometric qualities of the Dutch version of the PSC and whether it is suitable for and contributes to the early detection of psychosocial problems among Dutch children aged 7–12 by PCH. Results reveal a good internal consistency and validity using the CBCL as gold standard. However, lower cut-offs have to be used for Dutch children than for children from the USA because of the Dutch children's on-average lower scores.

Limitations

Methodological factors are unlikely to have affected these results. In general, the response rate was high (84%). Moreover, we used the CBCL as a criterion, which has been proven to be a valid measure for psychosocial problems. Because of complexity and high costs, structured clinical interviews such as the Diagnostic Interview Schedule for Children were not used as criteria.[23] Doing so may have provided additional information but differences with questionnaire-based information have been shown to be small.[24]

Fit with previous research on PSC and on other questionnaires used in PCH

This first study of the Dutch version of the PSC yielded results on reliability and on validity regarding the CBCL that are very similar to those found in comparable US samples. Jellinek et al. reported a sensitivity of 51.5% and a specificity of 95.4% at a cut-off of 0–27/28+, using the CBCL as a criterion in a sample of 206 children from the USA.[9,25] These values are very similar to those for a cutoff of 0–24/25+ for the Dutch version, i.e. 53.3% and 97.3% respectively (compare figure). We found the internal consistency of the PSC similar to that found by Jellinek et al.[25] but the results of our confirmatory factor analyses cast some doubt as to whether it measures a single latent, as did a previous study of Gardner et al.[26]

We found much lower mean scores on the PSC than have been found for comparable US samples. Mean scores on other symptom checklists such as the CBCL are also lower for Dutch children than for children from the USA.[27] Therefore, the Dutch children's lower mean PSC scores probably reflect real differences between these countries in the levels of symptoms reported by parents. This also implies that the cut-off for an elevated score on the PSC should be set lower for Dutch children than for children from the US. At this lower cut-off, the sensitivity and specificity of the Dutch version is similar to that of the US version. Moreover, the test characteristics of the PSC are comparable with or slightly better than those of most questionnaires currently used in Dutch PCH.[15,16,28]

Finally, we found higher mean PSC scores for boys and for children from single-parent families, similar to those found by Jellinek et al. for children in the USA.[10] We did not find differences in terms of parental education level, in contrast to the findings of Jellinek et al.,[10] but we did find elevated scores among children with unemployed parents, an indicator of familial socioeconomic status that was not studied by Jellinek et al.[10]

Implications

The results of our study imply that the PSC is useful for the early detection of psychosocial problems by PCH, especially if an adjusted cut-off is used. The PSC mostly detects behavioural and emotional problems, which are common in this age group. However, questions on more extreme behaviours such as the abuse of alcohol and drugs are not asked. Screening using the PSC is best carried out as a first step in a two-step process on the way to referral. A relatively low-cut-off can then be used to avoid missing too many cases. In a second step, cases flagged by the PSC should then be assessed by a CHP before making a final decision about referral. Parental responses show that some questions may require revision. In any event, the PSC is a useful aid for the early detection of psychosocial problems that could be considered for use in other countries as well.

References

1   Murthy RS, Bertolote JM, Epping-Jordan J, Funk M, Prentice T, Saraceno B, Saxena S. *The World health report 2001. Mental health: new understanding, new hope*. Geneva: World Health Organization, 2001.

2   Verhulst FC, van der Ende J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997; 36(7):901-909.

3   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001; 155(4):462-469.

4   Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems among toddlers in Dutch preventive child healthcare. *Arch Pediatr Adolesc Med* 2004; 158(8):811-817.

5   Moffitt TE, Caspi A. Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Dev Psychopathol* 2001; 13(2):355-375.

6   Durlak JA, Wells AM. Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *Am J Community Psychol* 1998; 26(5):775-802.

7   Nelson G, Westhues A, MacLeod J. A meta-analysis of longitudinal research on preschool prevention programs for children. *Prevention & Treatment* 2003; 6(31):1-32.

8   Murphy JM, Reede J, Jellinek MS, Bishop SJ. Screening for psychosocial dysfunction in inner-city children: further validation of the Pediatric Symptom checklist. *J Am Acad Child Adolesc Psychiatry* 1992; 31(6):1105-1111.

9   Jellinek MS, Murphy JM. Screening for psychosocial disorders in pediatric practice. *Am J Dis Child* 1988; 142(11):1153-1157.

10   Jellinek MS, Murphy JM, Little M, Pagano ME, Comer DM, Kelleher KJ. Use of the Pediatric Symptom Checklist to screen for psychosocial problems in pediatric primary care: a national feasibility study. *Arch Pediatr Adolesc Med* 1999; 153(3):254-260.

11   Pagano M, Murphy JM, Pedersen M, Mosbacher D, Crist-Whitzel J, Jordan P, Rodas C, Jellinek MS. Screening for psychosocial problems in 4-5-year-olds during routine EPSDT examinations: validity and reliability in a Mexican-American sample. *Clin Pediatr (Phila)* 1996; 35(3):139-146.

12   Jutte DP, Burgos A, Mendoza F, Ford CB, Huffman LC. Use of the Pediatric Symptom Checklist in a low-income, Mexican American population. *Arch Pediatr Adolesc Med* 2003; 157(12):1169-1176.

13   Stein BD, Jaycox LH, Kataoka SH, Wong M, Tu W, Elliott MN, Fink A. A mental health intervention for schoolchildren exposed to violence: a randomized controlled trial. *JAMA* 2003; 290(5):603-611.

14   Kelley ML, Fals-Stewart W. Psychiatric disorders of children living with drug-abusing, alcohol-abusing, and non-substance-abusing fathers. *J Am Acad Child Adolesc Psychiatry* 2004; 43(5):621-628.

15   Vogels T, Reijneveld SA, Brugman E, den Hollander-Gijsman ME, Verhulst FC, Verloove-Vanhorick SP. Detecting psychosocial problems among 5-6 years old children in Preventive Child Healthcare: The validity of a short questionnaire used in an assessment procedure for detecting psychosocial problems among children. *Eur J Public Health* 2003; 13:353-360.

16   Reijneveld SA, Vogels AG, Brugman E, Van Ede J, Verhulst FC, Verloove-Vanhorick SP. Early detection of psychosocial problems in adolescents: how useful is the Dutch short indicative questionnaire (KIVPA)? *Eur J Public Health* 2003; 13(2):152-159.

17   Word Medical Association Delaration of Helsinki: Ethical principles for medical research involving human subjects. [http://www.wma.net/e/policy/b3.htm].

18   Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile.* Burlington (VT): University of Vermont, Dept of Psychiatry, 1991.

19   Verhulst FC, van der Ende J, Koot HM. *Manual for the CBCL/4-18* [in Dutch]. Rotterdam: Erasmus University/Dept of Child and Adolescent Psychiatry, Sophia Childrens' Hospital, 1996.

20   Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46(12):1417-1432.

21   SPSS® for Windows®. [http://www.spss.com/spss].

22   Structural equation modeling to test telationships. [http://www.spss.com/amos/].

23   NIMH. *Diagnostic Interview Schedule for Children.* Rockville: National Institute of Mental Health, 1997.

24   Boyle MH, Offord DR, Racine YA, Szatmari P, Sanford M, Fleming JE. Adequacy of interviews vs checklists for classifying childhood psychiatric disorder based on parent reports. *Arch Gen Psychiatry* 1997; 54(9):793-799.

25   Jellinek MS, Murphy JM, Burns BJ. Brief psychosocial screening in outpatient pediatric practice. *J Pediatr* 1986; 109(2):371-378.

26   Gardner W, Kelleher KJ, Pajer KA. Multidimensional adaptive testing for mental health problems in primary care. *Med Care* 2002; 40(9):812-823.

27   Crijnen AA, Achenbach TM, Verhulst FC. Comparisons of problems reported by parents of children in 12 cultures: total problems, externalizing, and internalizing. *J Am Acad Child Adolesc Psychiatry* 1997; 36(9):1269-1277.

28   Vogels, AGC, Crone, MR, Hoekstra F, Reijneveld SA. *Three questionnaires to detect psychosocial problems among children aged 7-12* [in Dutch]. Leiden: TNO Quality of Life, 2005.

6 Comparing three short questionnaires to detect psychosocial dysfunction among primary school children: a randomized method

A.G.C. Vogels

M.R. Crone

F. Hoekstra

S.A. Reijneveld

Abstract

Good questionnaires are essential to improve the early identification of children with psychosocial dysfunction in community-based settings. Our aim was to assess which of three short questionnaires was most suitable for this identification among school-aged children.

A community-based sample of 2,066 parents of children aged 7-12 years (85% of those eligible) answered one of three questionnaires to be compared: the Strengths and Difficulties Questionnaire (SDQ), the Pediatric Symptom Checklist (PSC) and the PSYBOBA questionnaire. Which parent got which questionnaires was randomized. All parents answered the Child Behavior Checklist (CBCL), the main validity criterion. Preventive Child Healthcare (PCH) professionals assessed children's psychosocial functioning during routine health examinations. We assessed the scale structure, validity and usability of each questionnaire and determined whether the questionnaires could improve the identification based on clinical judgment.

Cronbach's α varied between 0.80 and 0.89. The Areas under the Receiver Operating Curve indices (criterion: the CBCL Total Problems Score (TPS)) were not significantly different. Sensitivities for a clinical TPS at a cut-off point with specificity ≥ 0.90 varied between 0.78 and 0.86 for the three questionnaires. Odds ratios, indicating information added by the questionnaires above clinical judgment were significant: 29.3 (PSC), 55.0 (SDQ) and 68.5 (PSYBOBA). Most PCH professionals preferred the SDQ for routine use in future.

We concluded that each questionnaire could improve the detection of psychosocial dysfunction among children substantially. Randomized comparative studies of questionnaires for early detection are a valuable method to obtain information for improving screening procedures.

6.1 Introduction

Psychosocial problems are quite common among children. Prevalence rates found vary with age, methods used and sample. Horowitz and co-workers reported that clinicians identified such problems among 27% of 1,886 4- to 8-year-old children who visited community-based, primary care pediatric practices.[1] In another study,[2] family physicians identified problems in approximately 22% of 898 children aged 5 to 15 years. Kelleher and co-workers reported that pediatric and family practice clinicians identified psychosocial problems among 19% of a national sample of children aged 4-15 years.[3,4] Only a minority of these children are treated for these problems. Verhulst[5] found that only 13% of children with behavioral or emotional problems were referred to mental healthcare. Reliable and valid questionnaires can improve early detection of such problems and the following treatment of these children, and thereby improve these children's prognosis significantly and substantively.[6]

Community pediatric services, like those in the USA and the Netherlands, offering routine healthcare services to the population as a whole, are in a unique position to detect children with psychosocial problems. In the Netherlands this early detection is an explicitly formulated task of the existing Preventive Child Healthcare (PCH). Without reliable and valid questionnaires, many children with problems are likely to be missed. For instance, Brugman and co-workers[7] showed that PCH, without such instruments, identified problems among 25% of the children, but 43% of the children with a clinical score on the CBCL were missed. Introducing valid instruments will also reduce the number of children without problems identified as having such problems

Several promising instruments are available and meet two essential conditions, set by the context of use in community services: they are short and easy to answer and to score. These are the Pediatric Symptom Checklist (PSC[8,9]), the Strengths and Difficulties questionnaire (SDQ[10,11]) and a newly developed Dutch instrument, the PSYBOBA (a Dutch acronym, standing for 'Questionnaire for psychosocial problems among primary school children aged 7 to 12').[12] These questionnaires ai84

m to detect behavioral and emotional problems in children in primary education and are available in parent form.

The aim of this study was to compare the psychometric properties and screening qualities of these questionnaires in a randomized diagnostic method to select the best questionnaire for routine use in community-based Preventive Child Healthcare (PCH) for primary school children. More specifically: which is the questionnaire that enables PCH to make the best distinction between children with and without problems, that offers PCH most information not already available from other sources, and that is most suitable in practical use, for parent and for PCH?

6.2 Methods

We developed a randomized procedure in which we compared the questionnaires on a number of predefined criteria, using data that we collected in an identical way for each

of the questionnaires to be assessed. To guarantee complete equivalence of data, we used a community sample of parents who all filled out the Child Behavior Checklist (CBCL), but were randomized to filling out one of three questionnaires to be evaluated (PSC, SDQ or PSYBOBA). Next, we obtained information from the PCH professionals, blinded for the parental questionnaire, about background characteristics and psychosocial problems detected by PCH professionals.

6.2.1 Sample
We obtained our sample in two steps. First, nine regional PCH centers, distributed all over the Netherlands, were found willing to participate in the study. Second, these PCH asked parents, invited for a regular check up of their child, to participate. We created three random subsamples, each receiving the CBCL and one of the questionnaires to be evaluated. Out of 2426 parents, 2066 were willing to participate (85%). Data in only 25 parental questionnaires were incomplete. Final response was 84%. We compared the respondents with the non-respondents on gender and age of the child, family situation (two-parent vs. one-parent family) and ethnicity (Dutch, from a country belonging to the Organisation for Economic Co-Operation and Development (OECD) (except Turkey) and other). We found some significant differences, but most of these were small effects only. The largest difference (Cohen's W=0.11) regarded ethnicity: more non-respondents originated from a non-OECD country (16% vs. 7%).

6.2.2 Randomization
Parents received one of the three questionnaires in a random way: the researchers put the questionnaires in closed envelopes and sent them to the PCH centers. The PCH centers sent these closed envelopes to the parents. Parents completed the questionnaires, put them in a closed envelope and gave them to PCH, which passed them to the researchers, again without opening the envelopes.

6.2.3 Power analysis of sample size
We aimed at a sample size of 700 respondents for each of the questionnaires to be evaluated. Earlier studies[13,14] showed that short questionnaires used in pediatric settings allow for an Area Under Curve Index (AUC) of about 0.90 with a clinical CBCL Total problem score as criterion. Subsample sizes of 700 are sufficient to detect, with α=0.05 and a power of 0.80, a difference between AUCs of 10 points around 0.90. PCH centers continued data collection until the required number of questionnaires was reached.

6.2.4 Procedure and measures
The data were collected during routine preventive health assessments of children aged to 12, between September 2003 and July 2004. Data collection procedures closely resembled those in other studies[7,13,15] and are known to result in high response rates. The CBCL and either the SDQ or the PSC or PSYBOBA were mailed to the parents along with the standard invitation to the preventive health assessment. These forms were returned in closed envelopes to the PCH professional, who sent them – unopened – to the researchers.

The PCH professionals then interviewed parent and child and examined the child. After each assessment the PCH professional answered the question: "Does the child have a psychosocial problem?" (yes, no). PCH was explicitly instructed to distinguish between risk indicators and emotional and behavioral problems as such, and when only risk indicators but no emotional or behavioral problems were detected, the child involved was to be coded as having no psychosocial problems.

The validated Dutch version of the CBCL was used to assess behavioral and emotional problems. The CBCL has a substantive concurrent and predictive validity, also in the Dutch version.[16,5] It allows for the calculation of a Total Problem Score (TPS) and several syndrome and broadband scores. Children were allocated to a normal, borderline or clinical range, using the 80th and 90th percentile of the TPS in the Dutch normative sample as cut-off points.[16]

The SDQ[10,17] is a brief behavioral screening questionnaire. It was developed originally in Great Britain, but is now available in more than 50 languages. Several studies indicate good psychometric properties, in different settings and in different nationalities.[18-22] It consists of 25 items and allows for the calculation of five subscales (Emotions, Behaviour, Peers, Hyperactivity and Prosocial Behaviour) and a total problem scale. Eight additional items, assessing the impact of problems, were not used in this study. We used the parent form of the Dutch version, available at WWW.SDQINFO.COM. The PSC is a 35 item parent questionnaire, assessing psychosocial dysfunction and developed in the USA.[23-25] Several studies have shown its good psychometric properties.[9,26,27] A single total problem score is calculated. Four additional items asking whether the child is treated or whether parents want any treatment, were not used in this study. No official Dutch version of the PSC was available. Therefore, this questionnaire was translated following a procedure advised by Guillemin,[28] using three independent translators and back-translators and the advice of the original authors. The PSYBOBA is a recently developed Dutch instrument, designed specifically for Dutch PCH.[12] It contains 26 items, on the child's behavior and emotions. These items allow for the calculation of a single problem score. Four additional items, on stressful life events and parental worries, are not used for the calculation of the problem score and were not included in the evaluation. We added three questions to each of the three questionnaires, asking how parents rated the length and the difficulty, and whether they had any remarks on the questionnaires.

Additionally, to get insight in the usability of the questionnaires, nine PCH professionals were asked to use each of the questionnaires in about 25 standard assessments. They rated the usability of the questionnaires, after ten assessments each, in a short rating list. This questionnaire contained ten items on how parents had answered the questionnaires, whether calculation of the scores was considered complicated and so on (see also Table 7). Having used each of the three questionnaires, the PCH professionals answered another questionnaire, in which they rated each questionnaires on a scale from 0 to 10 and indicated which questionnaire they preferred. Furthermore, parents were asked to rate the questionnaires on length and difficulty, and to report remarks they had on the questionnaires.

The study was approved by the local Medical Ethical Committee.

6.2.5 Analyses

We compared the questionnaires on a number of criteria, developed in studies published earlier.[13,14] The main criterion variable was a clinical CBCL TPS. As PCH aims to detect problems in an early stage, additional analyses were done using a CBCL TPS score in the borderline range.

First, we assessed whether the three subsamples were comparable in terms of background characteristics and criterion variable. Differences were tested with bi-variate Chi^2-tests.

Next, we assessed the scale structure of the questionnaires. Cronbach's α was calculated. We determined the fit between the scale structure and the observed data using Amos 5 Structural Equation Modeling (SEM).[29] We used SEM in stead of the more usual explorative factor analytical techniques, as SEM allows for testing the fit between the data with the scale structure. The models tested were considered as fitting when the Parsimony Corrected Fit Index (PCFI) was higher than 0.90.

Third, we assessed the validity of the questionnaires, using the CBCL TPS as criterion. Validity in this context refers to the extent to which the questionnaires can distinguish between children with and without problems. Pearson correlation coefficients and kappas were calculated. We calculated the AUCs and tested differences between them by means of bootstrapping. Furthermore, sensitivity and specificity of the three questionnaires were calculated. The cut-off points suggested by the original authors resulted in considerable differences in the prevalence of elevated scores (10% for the SDQ, 5% for the PSC and 32% for the PSYBOBA) and therefore in highly different sensitivities and specificities, while the Area Under Curve indices were comparable. For that reason, we defined new cut-off points, which led to a specificity in our data set of at least 0.90, and used these cut-off point in the analyses.

Fourth, we assessed the added value of the instruments: to what extent do they offer new information, compared to other readily available information that may indicate possible problems. We performed logistic regression analyses to predict a clinical and borderline CBCL TPS, using demographic risk indicators as predictors. Then we added the elevated scores to the model and checked whether this improved the prediction. This is comparable to what we did in two earlier studies.[13,14] We extended these analyses and also used the question whether the PCH professional had detected any problem (yes or no) as a predictor in the model, before entering the elevated scores. The height of the odds ratio for elevated scores was used to measure the added value each questionnaire offered. Where relevant, differences between betas were tested using the formula $(\beta1-\beta2)/(se1-se2)0.5 > 1.96$.

Finally, the usability of the questionnaire in daily practice was assessed. For parents, differences in response rate, item non-response, and opinions on difficulty and length and the number of critical remarks were tested with ANOVA and Chi^2-tests. For the PCH professional's opinion on usability we counted the number of optimal ratings over the individual health assessments for which each questionnaire was used. The comparative ratings of the questionnaires by PCH professionals were inspected and described. No tests were done on these data, as only nine PCH professionals participated in the pilot. Unless otherwise specified, analyses were done in SPSS 11.5 or 12.

6.3 Results

6.3.1 Differences between subsamples

Table 1 presents information on demographic characteristics and the main criterion variable for the three subsamples. No statistically significant differences between the subsamples were found.

Table 1   Characteristics of the respondents, by subsample defined by questionnaire filled in.

	PSYBOBA (n=660) %	PSC (n=674) %	SDQ (n=770) %	Total (n=2041) %
Child's gender				
Male	51	49	49	50
Female	49	51	51	50
Child's age				
7-9 years	47	48	46	47
10-12 years	53	52	54	53
Ethnic background				
Dutch	84	82	83	83
non-dutch	6	7	7	6
Unknown	10	10	10	10
Family composition				
two parents	85	88	86	86
one parent	9	8	10	9
Other	5	4	4	5
Parental employment				
no paid job	3	3	3	3
two parents with paid job	52	54	53	53
one parent with paid job	35	34	34	34
Unknown	9	9	10	9
Parental highest completed education				
none or only primary (max. 8 yrs)	2	3	3	3
lower vocational (max. 12 yrs)	24	27	25	25
higher vocational (max 16 yrs)	33	29	32	31
University / higher professional (min. 17 yrs)	35	35	33	35
Unknown	6	5	6	6
CBCL TPS score in clinical range	10	9	8	9
No (% of final response) of children with a clinical TPS	66(10%)	60(9%)	56(8%)	182(9%)

Table 2   Response by questionnaire filled in

	PSYBOBA	PSC	SDQ	Total
No of parents invited to participate	796	815	814	2425*
No (%) of parents refusing	113 (14%)	103 (13%)	78 (10%)	294 (12%)
No (%) of parents not returning the questionnaires	15 (2%)	25 (3%)	25 (3%)	65 (3%)
No (%) of parents with incomplete data	8 (1%)	13 (2%)	4 (0%)	25 (1%)
Final response:				
No (%) of parents included in the analyses	660 (83%)	674 (83%)	707 (87%)	2041 (84%)
* In one case it is not known which questionnaires the parents received.				

Table 2 presents figures on response rates by questionnaire. Final response is largest for the SDQ (Chi^2 =22.5; df=1,2; p < .05). The percentage of children with a clinical CBCL TPS in the final response did not differ significantly between the three subsamples.

6.3.2 Scale structure

The internal consistencies of the total problem scales of the three questionnaires were satisfactory: 0.80 for the SDQ, 0.87 for the PSYBOBA and 0.89 for the PSC. The differences in Cronbach's α could be completely explained by the differences in number of items: application of the Spearman Brown Prophecy Formula resulted in exactly the same corrected α's. Internal consistencies for the SDQ's subscales varied between 0.55 and 0.78.

Structural equation modeling showed a poor fit of the single-scale models with the data (SDQ: PCFI=0.66; PSC: PCFI=0.66; PSYBOBA: PCFI=0.65). For the SDQ a more subtle model was evaluated, reflecting the questionnaire's subscales. This model had to be rejected, too (PCFI=0.74).

6.3.3 Validity

Table 3 presents data on the validity of the three questionnaires, using a clinical CBCL TPS and a borderline or clinical CBCL TPS as criterion, for all children and for children who were never treated for any psychosocial problem in the past. The Table also presents the cut-off points used, when calculating kappa and sensitivity and specificity. Differences between the three questionnaires were small. We repeated the analyses for cut-off points which would result in a specificity of at least 0.95. This cut-off point would result in a sensitivity for a clinical CBCL TPS varying between 0.73 (for the SDQ) and 0.62 (for the PSYBOBA). Again, no significant differences were found.

Exclusion of children who had been under treatment resulted in almost identical sensitivities and specificities. Use of a borderline TPS as the criterion resulted in somewhat lower sensitivities.

6.3.4 Added value

Three demographic variables showed a significant association with a clinical CBCL TPS: gender, country of origin (OECD country versus other) and family composition (two biological parents, one biological parent, no biological parents). These variables were included as possible predictors of a clinical or borderline/clinical CBCL TPS. Adding elevated scores to the model resulted in a significant (p<.001) and substantial improvement of the model, for each of the three questionnaires (Table 4).

The differences between the calculated ORs were small, in relation to the 95% confidence intervals (CI). In the analysis using all children, the OR of the PSC lies just below the 95% CI of the OR of the PSYBOBA. The difference between the two betas was not significant, however (p=0.11).

Table 5 presents the results of the extended added value analyses, in which problems as detected by the PCH professionals were also taken into account. Overall, PCH professionals identified 27% as having some psychosocial problem. The ORs reported in Table 5 are lower than those in Table 3, which suggests that part of what is detected

by the questionnaires, was also seen by the PCH professionals without the aid of questionnaires. Yet, in all analyses adding the elevated scores to the model resulted in a very significant (p < .001) improvement of the prediction. The differences between the questionnaires were not significant.

Table 3 *Validity indicators, using clinical and borderline CBCL TPS as criteria: Pearson's r, Kappa, sensitivity and specificity; for all children and only for children not being treated*

	PSYBOBA	PSC	SDQ
Pearson's r	.81	.81	.77
Kappa (cut-off point used)	.59 (>14)	.52 (>20)	.53 (>11)
AUC (95% CI)	0.96 (0.94 - 0.98)	0.93 (0.92 - 0.96)	0.95 (0.93 - 0.98)
Clinical TPS			
All children			
Cut-off point	>14	> 20	> 11
Sensitivity (95% CI)	0.86 (0.78 - 0.94)	0.78 (0.68 – 0.88)	0.86 (0.77- 0.95)
Specificity (95% CI)	0.91 (0.89 – 0.93)	0.92(0.89 – 0.93)	0.90 (0.88 – 0.92)
Positive Predictive value	0.51 (0.41 – 0.61)	0.45 (0.35 – 0.55)	0.43 (0.34 – 0.52)
Children currently not under treatment			
Cut-off point	> 14	>19	>11
Sensitivity (95% CI)	0.82 (0.71 – 0.93)	0.81 (0.69-0.93)	0.80 (0.67 – 0.93)
Specificity (95% CI)	0.92 (0.91 - 0.95)	0.91 (0.89 – 0.93)	0.92 (0.90 - 0.94)
Positive Predictive value	0.44 (0.33 – 0.55)	0.40 (0.30 – 0.51)	0.37 (27 – 48)
Borderline TPS			
All children			
Cut-off point	>13	>18	>10
Sensitivity (95% CI)	0.73 (0.64 – 0.82)	0.72 (0.64 - 0.80)	0.79 (0.71 – 0.87)
Specificity (95% CI)	0.91 (0.89 - 0.93)	0.91 (0.89 – 0.93)	0.90 (0.88 – 0.92)
Positive Predictive value	0.58 (0.49 – 0.67)	0.62 (0.53 – 0.70)	0.59 (0.49 – 0.68)
Children currently not under treatment			
Cut-off point	>13	> 17	>10
Sensitivity (95% CI)	0.66 (0.55 – 0.77)	0.74 (0.65 - 0.83)	0.75 (0.65 – 0.85)
Specificity (95% CI)	0.91 (0.89 – 0.93)	0.91 (0.89 – 0.93)	0.90 (0.89 - 0.93)
Positive Predictive value	0.48 (0.38 – 0.58)	0.57 (0.47 – 0.66)	0.49 (0.39 - 0.59)

Table 4 *Results of the first added value analysis: adjusted[*] odds ratios for elevated scores on the three questionnaires for a CBCL TPS score in the clinical and borderline clinical range*

	PSYBOBA OR (95% CI)	PSC OR (95% CI)	SDQ OR (95% CI)
Criterion: Clinical TPS			
All children			
Elevated score yes (versus no)	89.3 (38.0 – 210.6)	35.6 (17.8 – 71.4)	71.7 (30.7 – 167.4)
Children currently not under treatment			
Elevated score yes (versus no)	71.8 (28.4 – 181.1)	46.3 (19.6 – 109.5)	52.9 (22.0 – 127.5)
Criterion: Borderline TPS			
All children			
Elevated score yes (versus no)	27.7 (16.0 – 48.0)	27.1 (16.0 – 45.8)	40.6 (22.1 – 74.8)
Children currently not under treatment			
Elevated score yes (versus no)	21.0 (11.5 – 38.4)	29.2 (162. – 52.7)	34.8 (18.0 – 67.1)
* adjusted for the effects of gender, country of origin and family composition			

Table 5 *Results of the extended added value analyses: adjusted[*] odds ratios for elevated scores on the three questionnaires for a CBCL TPS score in the clinical and borderline clinical range*

	PSYBOBA OR (95% CI)	PSC OR (95% CI)	SDQ OR (95% CI)
Criterion: Clinical TPS			
All children			
Elevated score yes (versus no)	68.5 (28.3 – 165.6)	29.3 (14.4 – 59.8)	55.0 (23.1 – 131.3)
Children currently not under treatment			
Elevated score yes (versus no)	65.4 (24.8 – 172.4)	40.1 (16.7 – 96.3)	44.2 (18.0 – 108.3)
Criterion: Borderline TPS			
All children			
Elevated score yes (versus no)	22.1 (12.4 – 39.3)	22.8 (13.3 – 39.0)	31.3 (16.8 – 58.6)
Children currently not under treatment			
Elevated score yes (versus no)	19.0 (10.1 – 35.9)	26.8 (14.7 – 49.0)	29.6 (15.1 -58.6)
* adjusted for the effects of gender, country of origin and family composition and for problems identified by PCH			

6.3.5 Usability – parents

The response rate for the SDQ (87%) was slightly higher than for the PSYBOBA and the PSC (both 84%) (Chi2=22.5, df=2, p <.01). The PSC showed the highest mean number of unanswered questions (n=0.26), compared to 0.06 for the PSYBOBA and 0.05 for the SDQ (F=7.4, df=2, 2076, p< 0.001).

Parents' ratings of length and difficulty of the questionnaires showed significant but small differences between the questionnaires. Twelve percent found the PSYBOBA long or too long, compared to 20% for the PSC and 19% for the SDQ. The percentage of parents rating the questionnaires as difficult varied between 41% % (PSC) and 31% (PSYBOBA) Finally, 20% of the parents made a critical remark on the PSC, compared to 9% for the PSYBOBA and 10% for the SDQ. They criticized the ambiguity of some questions (*e.g.* "Spends more time alone …" without specification of how to compare) and the discrepancy between items and answering categories (*e.g.* "School grades dropping" to be answered with never, sometimes, often).

6.3.6 Usability – PCH professionals

PCH professionals rated the three questionnaires on 9 aspects after using them during about 76 examinations. The mean number of optimal ratings was highest for the PSC (6.0 out of 9). For the PSYBOBA it was 5.6 and for the SDQ 4.5. This difference is significant (F=25.2, df=2, 241, p< .0001). The largest differences were found on items referring to the complexity of the calculation of scores.

After completing the routine examinations with all three questionnaires the PCH professionals rated the questionnaires on a scale from 0 to 10. Mean rating for the SDQ and the PSYBOBA was 6.3, higher than that for the PSC (5.5). When asked which questionnaire they would like to use in the future, four of the eight PCH professionals chose the SDQ, two the PSYBOBA and only one the PSC. One PCH professional did not make a choice.

Table 6   *Percentage of examinations resulting in an optimal rating by PCH professionals on nine questions on usability by questionnaire*

	PSYBOBA	PSC	SDQ
	%	%	%
All items answered	97	88	87
Not irritating at all for the parent	53	57	62
Fully understood	65	62	69
Questionnaire led parents to reflection of the child's mental health	10	8	9
Calculation of scales scores: not difficult	75	98	29
Calculation of scale scores not time consuming	71	88	11
Conversation, based on questionnaire, with parents was useful	11	11	18
Questionnaire covered all parents perceived problems	80	90	87
Questionnaire covered all problems perceived by PCH professional	78	84	78
Mean no. of optimal ratings (st. dev)	5.6 (1.6)	4.5 (1.5)	6.0 (1.1)
No of exams	70	81	78

6.4 Discussion

This study compared three questionnaires (SDQ, PSC and PSYBOBA) in order to decide which one was most suited to improve identification of psychosocial problems among children aged 7 to 12 in community health services. The internal consistency of the overall scales was high. All questionnaires had a satisfactory sensitivity, at a specificity of 0.90, for problems defined as a clinical CBCL TPS score, and a somewhat lower sensitivity for problems defined as a borderline CBCL TPS score. All three questionnaires offered substantial added value, improving the identification of children with problems to a situation in which a PCH professional has to rely on readily available health indicators and /or the routine examination.

The PSC resulted in a higher item non response, and one in five parents made some critical remark about the PSC. Due to the simplicity of score calculation, it was rated more favorably by nine PCH professionals, in daily practice. Yet, in an overall rating by these professionals the PSC was rated less favorably than the SDQ and PSYBOBA.

In this study data were collected using a methodology that closely resembles the way PCH works. This improves the external validity of our results. However, one caveat is important: the main aim during sampling was to guarantee similarity between the three subsamples, not an overall representativeness for the Dutch population. This resulted, unfortunately, in a clear underrepresentation of ethnic minorities. There is evidence that the psychometric performance of questionnaires differs between different cultures. The results therefore need confirmation among ethnic minorities.

We used the CBCL TPS as criterion. Although the CBCL is one of the best instruments available and is often used for evaluation purposes , it cannot be regarded as the ultimate golden standard. It underestimates, for example, the prevalence of internalizing problems, when compared to instruments answered by children themselves. Other studies used psychiatric interviews or assessments by mental health professionals as standard.[25] Due to financial limitations this was not possible in this study.

The absence of an ultimate golden standard means that the results of the study should be interpreted carefully. The high convergence between the three questionnaires and the

CBCL indicates that they can be used as a valuable tool, supporting the professional's assessment. Yet, it seems unwise to use these short questionnaires as pre-selection tools, limiting further assessment to those children with elevated scores. The risk of missing serious problems would be too great.

Our data on the validity and reliability of the SDQ and the PSC are comparable to those published by other authors.[26,27,30,31] We found, however, no other studies which question the scale structures of the questionnaires. Probably, our choice for the rigorous SEM as analytical tool instead of the more usual factor analytical approaches is the key factor. The combination of high internal consistencies and the negative SEM results may come as a surprise. What the SEM analyses showed, however, is that the concepts, as implied by the (sub)scale scores, despite the internal consistencies, are an inadequate description of the way the items are related to each other. In other words: the items provide information not covered by the scale scores. Healthcare providers should, therefore, not rely on the scale scores alone, but also carefully check the answers on individual items.

The percentage of incompletely filled out PSCs was rather high, compared to figures reported elsewhere.[9] This may be related to the critical remarks made by parents on the ambiguity of items and the lack of fit between item and answering categories. These problems cannot be attributed to the Dutch translation, as the same remarks can be made about the original PSC.

An essential element of the design of this study is that it used randomization and aimed at a comparison of three questionnaires. We know of no other studies that used a similar design. Our approach is comparable to what is now rapidly becoming standard in studies assessing effectiveness and economic evaluations of interventions. Such studies do not try to assess the effectiveness or costs as such, but compare specific interventions with other interventions or usual care. Such an approach is far more helpful in guiding health policy decisions. We feel that such a comparative approach is worthwhile, too, in the evaluation of questionnaires to be used in healthcare. Only a systematic comparison can guarantee that the best instrument available will indeed be chosen.

Few studies assessed the added value of using questionnaires, as we did. The only studies to compare our results with are our own evaluations of two other questionnaires, that assessed the added value of questionnaires, as compared to risk indicators.[13,14] In this study we extended the added value analyses, by including the signals detected by PCH professionals during routine examinations into the analysis. These extended analyses give a better indication of the real added value, as they compare the quality of questionnaire-based detection to what is now standard practice for this age group, at least in the Dutch healthcare system. We used logistic regression for these analyses. Pepe et al.[32] point to limitations of using ORs in gauging the performance of screening markers. The ORs we found, however, were well above 16, the value they consider as the minimum to be of any relevance.

One question is open to debate: what are the best cut-off points to be used? In general, there is no single optimal cut-off point. Jutte[33] for example, proposed to use a cut-off point of 12 for the PSC in a low-income Mexican population, far lower than the cut-off points proposed originally. Which cut-off points to use depends on several factors, such as the seriousness of the problems that should be detected, the sensitivity and specificity

needed, the prevalence of problems to be detected, available resources (for further assessment and treatment or referral) and the specific population targeted. Based on our results we would like to suggest Dutch PHC to use the cut-off points presented in Table 3, in order to be able to distinguish between low, suspect and elevated scores. These suggested cut-off points are different from those suggested by Goedhart et al.[21] or van Ede et al.[12] We feel, however, that decisions on cut-off points in screening type procedures should be based on content validity and not only on considerations related to the statistical distribution of scores.

Which of the three questionnaires, then, would be the best choice? The PSC's sensitivity was somewhat less than that for the other questionnaires. The added value of the PSC in detecting children with clinical CBCL TPS was relatively low. One in five of the parents made critical remarks on the PSC and item non response was also higher. In the Netherlands, the PSC would therefore be an unlikely choice. The psychometric performance of the PSYBOBA and the SDQ were similar. Although more PCH professionals preferred the SDQ as the instrument to use in the future, it was rated less positively in practical use, mainly because of the relative complexity of calculating the (sub)scale scores. We found little support for the supposed scale structure and showed that the SDQ's total problem score is a strong indicator of problems. When the primary aim is to make a first distinction between children who probably have problems that need attention and those who do not, the SDQ's single Total Problem scale is a sufficient indicator. This would make using the SDQ far more simple.

So, psychometric performance and user friendliness do not offer conclusive arguments for a choice for either the PSYBOBA or the SDQ. Other arguments can therefore be taken into considerations, for example which instrument is accepted most by other professionals to whom children may be referred to by PCH or in scientific research.

References

1   Horowitz LM, Wang PS, Koocher GP et al. Detecting suicide risk in a pediatric emergency department: development of a brief screening tool. *Pediatrics* 2001 May;107(5):1133-7.
2   Wildman BG, Kinsman AM, Logue E, Dickey DJ, Smucker WD. Presentation and management of childhood psychosocial problems. *J Fam Pract* 1997 January;44(1):77-84.
3   Moffitt TE, Caspi A. Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology* 2001;(13):355-75.
4   Verhulst FC, Dekker MC, Ende van der J. Parent, teacher and self-reports as predictors of signs of disturbance in adolescents: whose information carries the most weight? *Acta Psychiatr Scand* 1997 July;96(1):75-81.
5   Verhulst FC, Ende van der J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997 July; 36(7):901-9.

6   Durlak JA, Wells AM. Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *Am J Community Psychol* 1998 October;26(5):775-802.

7   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001 April;155(4):462-9.

8   Jellinek MS, Murphy JM. Use of the Pediatric Symptom Checklist in outpatient practice. *Curr Probl Pediatr* 1990;602-9.

9   Jellinek MS, Murphy JM, Little M, Pagano ME, Comer DM, Kelleher KJ. Use of the Pediatric Symptom Checklist to screen for psychosocial problems in pediatric primary care: a national feasibility study. *Arch Pediatr Adolesc Med* 1999 March;153(3):254-60.

10   Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 2000 December;177:534-9.

11   Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *J Abnorm Child Psychol* 1999 February;27(1):17-24.

12   Ede van J, Rover de C, Verhappen MJ, Vries-Lequin Id. *PSYBOBA. A pilot study on the usability of a questionnaire for psychosocial problems among primary school children aged 7 till 12 [in Dutch]*. GGD Nederland; 2003.

13   Vogels T, Reijneveld SA, Brugman E, Hollander-Gijsman M, Verhulst FC, Verloove-Vanhorick SP. Detecting psychosocial problems among 5-6-year-old children in preventive Child Healthcare: the validity of a short questionnaire used in an assessment procedure for detecting psychosocial problems among children. *Eur J Public Health* 2003 December;13(4):353-60.

14   Reijneveld SA, Vogels AG, Brugman E, Ede van J, Verhulst FC, Verloove-Vanhorick SP. Early detection of psychosocial problems in adolescents: how useful is the Dutch short indicative questionnaire (KIVPA)? *Eur J Public Health* 2003 June;13(2):152-9.

15   Zeijl E, Crone MR, Wiefferink CH, Keizenkamp S., Reijneveld SA. *Children in the Netherlands [in Dutch]*. The Hague: SCP; 2005.

16   Verhulst FC, Koot HM. *Manual for the CBCL/4-18 [in Dutch]*. Rotterdam: EUR/Sophia Kinderziekenhuis; 1996.

17   Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 1997 July;38(5):581-6.

18   Mathai J, Anderson P, Bourne A. The Strenghts and Difficulties Questionnaire (SDQ) as a screening measure prior to admission to a Child and Adolescent Mental Health Service (CAMHS). *Aust e-J Advancement Mental Health* 2001;1(3):1-12.

19   Koskelainen M, Sourander A, Kaljonen A. The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *Eur Child Adolesc Psychiatry* 2000 December;9(4):277-84.

20   Klasen H, Woerner W, Rothenberger A, Goodman R. [German version of the Strength and Difficulties Questionnaire (SDQ-German)--overview and evaluation

of initial validation and normative results]. *Prax Kinderpsychol Kinderpsychiatr* 2003 September;52(7):491-502.

21   Goedhart A, Treffers F, Widenfelt van B. Inquiring about mental health problems among children and adolescents: the Strenghts and Difficulties Questionnaire (SDQ) (in Dutch). *Maandbl Geestelijke Volksgezondh* 2003;58:1018-35.

22   Widenfelt van BM, Goedhart AW, Treffers PD, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003 December;12(6):281-9.

23   Anderson DL, Spratt EG, Macias MM et al. Use of the pediatric symptom checklist in the pediatric neurology population. *Pediatr Neurol* 1999 February;20(2):116-20.

24   Jellinek M, Evans N, Knight RB. Use of a behavior checklist on a pediatric inpatient unit. *J Pediatr* 1979 January;94(1):156-8.

25   Jellinek M, Murphy JM, Robinson J, Feins A, Lamb S, Fenton T. Pediatric Symptom Checklist: screening schoolage children for psychosocial dysfunction. *J Pediatr* 1988;112(2):201-9.

26   Jellinek MS, Murphy JM. Screening for psychosocial disorders in pediatric practice. *Am J Dis Child* 1988 November;142(11):1153-7.

27   Jellinek MS, Murphy JM, Burns BJ. Brief psychosocial screening in outpatient pediatric practice. *J Pediatr* 1986 August;109(2):371-8.

28   Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993 December;46(12):1417-32.

29   Byrne BM. *Structural Equation Modeling with AMOS. Basic Concepts, applications and programnming*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc. Publishers; 2001.

30   Muris P, Meesters C, van den BF. The Strengths and Difficulties Questionnaire (SDQ)--further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *Eur Child Adolesc Psychiatry* 2003 January;12(1):1-8.

31   Widenfelt van BM, Goedhart AW, Treffers PD, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003 December;12(6):281-9.

32   Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004 May 1;159(9):882-90.

33   Jutte DP, Burgos A, Mendoza F, Ford CB, Huffman LC. Use of the Pediatric Symptom Checklist in a low-income, Mexican American population. *Arch Pediatr Adolesc Med* 2003 December;157(12):1169-76.

7 A comparison of four classification methods based on the parent-rated Strengths and Difficulties Questionnaire as used in the Dutch Preventive Child Healthcare system

M.R. Crone

A.G.C. Vogels

F. Hoekstra

P.D.A. Treffers

S.A. Reijneveld

Abstract

Background:

Validated questionnaires can support the identification of psychosocial problems by Preventive Child Healthcare (PCH) system. This study assesses the validity and added value of four scoring methods used with the Strengths and Difficulties Questionnaire (SDQ) for the identification of psychosocial problems among children aged 7-12 by the PCH.

Methods:

We included 711 children (response 87%) aged 7–12 undergoing routine health assessments in nine PCH services across the Netherlands. Child health professionals interviewed and examined children and parents. Prior to the interview, parents completed the SDQ and the Child Behavior Checklist (CBCL), which were not shown to the professionals. The CBCL and data on the child's current treatment status were used as criteria for the validity of the SDQ. We used four SDQ scoring approaches: 1) an elevated SDQ Total Difficulties Score (TDS), 2) parent-defined difficulties, 3) an elevated score for emotional symptoms, behavior problems or hyperactivity in combination with a high impairment score, and 4) a combined score: an elevated score for any of these three methods.

Results:

The areas under the ROC curve ranged from 0.68 to 0.86 for the four scoring methods, generally indicating good validity. All four methods added significantly to the identification of problems by the PCH. Classification based on the TDS yielded results similar to more complicated methods.

Conclusion:

The SDQ is a valid tool for the identification of psychosocial problems by PCH. As a first step, the use of a simple classification based on the SDQ TDS is recommended.

7.1 Introduction

Psychosocial problems, such as behavioral, emotional, and educational problems, are very prevalent among children and adolescents, and may interfere severely with their everyday functioning. Only a minority of the children with such problems receive mental healthcare. In a study of more than 2,000 Dutch children, only 13% of the children with behavioral and emotional problems had been referred to mental health services in the year prior to the assessment.[1] Early treatment, however, may reduce these problems, if they are accurately identified.[2]

In the Netherlands, the Preventive Child Healthcare (PCH) system is one of the most important low-threshold services for the early identification of emotional and behavioral problems in children. Physicians and nurses working in the PCH routinely offer preventive healthcare to all children aged 0-19 living in the Netherlands. More than 90% of all children undergo three to four assessments by a child health doctor or nurse during their school careers, in both primary and secondary school.[3,4] In the Netherlands, municipalities are obliged by law to guarantee proper access to this type of care, free of charge, including the early identification of psychosocial problems.

However, several studies have shown that when the PCH does not use validated questionnaires, only half of the children with emotional or conduct problems are identified.[3,4,5] Validated questionnaires may help in the identification of these problems by PCH.[6,7] For children aged 7-12 years, however, there is no short validated questionnaire for use by the PCH.

The Strengths and Difficulties Questionnaire (SDQ) is a promising option in this respect. It was developed by Robert Goodman to support the early identification of conduct and emotional problems.[8,9] It is a brief measure covering the most important current domains of child psychopathology (i.e. emotional symptoms, conduct problems, hyperactivity-inattention, and peer problems) that can be completed by parents, teachers and young people themselves. The SDQ Parent Form, that we used in our study, consists of 25 symptom items, one item relating to the severity of problems as perceived by parents and seven items assessing difficulties in functioning associated to the reported problems. The psychometric properties and validity of the SDQ have already been shown to be good in a number of countries[10,11,12,13,14] including the Netherlands.[15,16] However, its appropriateness and added value for use by the PCH has not yet been assessed.

Bourdon et al.[17] used four SDQ scoring methods in a US setting to identify children who may have serious mental health difficulties. Their approach was based on the three components of the SDQ (symptom items, severity as perceived by the parents and impairment in functioning (see Method section for details)). The percentages of children identified varied according to the scoring method. Using service contact/use for a mental health reason as validation criterion, they found highly significant associations between service contact/use and each scoring method. The scoring method using parent-defined severity identified the highest percentage of children with a service contact/use. Bourdon et al.[17] therefore conclude that parental judgment of the severity of children's difficulties may be a key indicator in bringing those difficulties to the attention of general medical and mental health professionals.

In our view, contact with and use of mental health services is of limited value as a measure for the validation of questionnaires such as the SDQ. Research has shown that many children with serious problems are not referred to such services.[1] If this variable is used as the main criterion, the children with problems who have no contact with mental health services will not be identified. Bourdon et al.[17] did not have data relating to a validated overall instrument on emotional and conduct problems, such as the Child Behavior Checklist (CBCL), to validate the scoring methods. This study assesses the validity and added value of the four classification methods of the SDQ for the detection of emotional and conduct problems by the PCH, using both the CBCL and current treatment for psychosocial problems as criteria for validity.

7.2 Methods

7.2.1 Population

We obtained our sample using a two-step procedure. In the first step we selected a sample of PCH services. These PCH services then collected data relating to children aged 7-12 years. Thirty-five child health professionals, from nine PCH services, participated in this study. A total 814 parents and their children were asked to participate in this study; 10% refused to participate and 3% did not return the questionnaire, resulting in a response of 711 children (87%). Respondents were representative of the total sample in terms of age and gender, but non-response was higher for children of immigrant/minority origin (27.4% vs. 12.2% for children from Dutch origin).

7.2.2 Measures and procedures

Data were obtained during routine health assessments. Before the assessments parents filled out the CBCL and the SDQ. The parents gave both questionnaires to the child health professional, who passed them on to the researchers without opening them. The child health professionals interviewed the children and parents about mental health and background and examined the children with the help of a structured questionnaire including questions on life-events and current treatment for psychosocial problems. After each assessment, the health professional answered the following questions: 'Is the child currently being treated for psychosocial problems?' and 'Does the child have a psychosocial problem, at present?' (yes, no) and scored the severity (mild, moderate or severe) and type of problem(s) identified, using a pre-coded list.

In this study we used the parent version of the SDQ 4-16.[8,15,16] The questionnaire consists of 25 symptom items describing positive and negative attributes of children and adolescents that can be allocated to 5 subscales of 5 items each: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and pro-social behavior. Each item has to be scored on a 3-point scale: 0='not true', 1='somewhat true', and 2='certainly true'. A SDQ Total Difficulties Score (TDS) can be calculated by aggregating the scores for the emotional symptoms, conduct problems, hyperactivity-inattention, and peer problems subscales (range 0-40). The SDQ also contains an impact

supplement that asks the parents about the severity of the problems as perceived by the parents and enquires about duration, distress, social impairment, and burden for others. A three-point scale is used for each item: 0=not at all/only a little, 1=quite a lot, 2=a great deal. An impairment score was calculated by aggregating the scores for distress and social impairment.[15,16]

We dichotomized the CBCL and SDQ scale scores for the analyses. For the CBCL, we used the standard Dutch cut-off points for dichotomizing.[18] Dutch children tend to score lower on the SDQ than UK children; in the Netherlands about 6% of all children score above the UK cut-off point ($\geq$ 17). Use of this cut-off point would have led to low sensitivity indices (0.52 for a clinical CBCL score and 0.27 for currently being treated). Therefore, we also computed sensitivity and specificity at a cut-off point that yielded a prevalence rate similar to that in the UK (10%). The most appropriate cut-off therefore was a SDQ TDS of 14 and higher.

Bourdon et al.[17] developed four scoring methods to identify children who may have serious mental health difficulties. These methods were based on the three components of the SDQ (SDQ TDS, parent-reported severity and impairment in functioning). Bourdon et al. classified children as having problems in four ways:

1. children with a score on the SDQ TDS above the cut-off point
2. children whose parents reported definite or severe difficulties on the impact supplement of the SDQ
3. children with scores above the UK cut-off point for emotional symptoms, conduct problems, or hyperactivity-inattention in combination with an impairment score above the cut-off point
4. combination: children classified as having problems using any of the first 3 methods.

These four classification methods were included in the analyses.

The CBCL was used as a gold standard for parent-reported conduct and emotional problems during the preceding six months.[18] The reliability and validity of the CBCL established by Achenbach were confirmed for the Dutch translation.[19] The CBCL consists of 20 competence items and 120 problem items. We used only the problem items. Parents indicated the presence of problems, choosing from three categories (no problem, sometimes/a little, often/a lot). We computed two broadband groups of syndromes – Internalizing and Externalizing – and a Total Problems Score (TPS). Cases were subsequently allocated to a normal or a clinical range in accordance with the scoring distributions in the Dutch normative sample.[18]

Child and family background characteristics assessed by the CHP were: gender, age, ethnicity, family characteristics (number of parents), income, educational level of the mother and employment status of the parent(s). Ethnicity was based on the native country of both biological parents. The country was coded as non-industrialized if at least one parent was born outside a member country of the Organisation for Economic Co-operation and Development. Turkey, however, was classified as a non-industrialized country, too.

7.2.3 Analysis

The analyses assessed the validity of the four classification methods and their added value for the identification of children with problems by the PCH. The validity of the different scoring methods was assessed using Receiver Operating Characteristics analyses (ROC), with currently being treated for psychosocial problems and the dichotomized (normal/clinical) CBCL Total Problem and Internalizing and Externalizing scores as criteria. The resulting Area Under Curve (AUC) and sensitivity and specificity indices will be presented.

We then determined the added value of the four classification methods, i.e. we assessed to what extent each of the four methods contributes to the distinction between children with and without problems, after taking into account the identification by PCH, based on clinical judgment after the standard health examination. To this end, we first performed logistic regression analysis with each of the criterion measures as the dependent variable and the identification by PCH as the independent variable. We then added the SDQ-based classification methods. The significance of the change in the log likelihood ratio of these second models was used to determine whether adding the classification methods contributed to a better distinction. We will present the resulting odds ratios (OR) of the second models, when the change in log likelihood is significant.

7.3 Results

The characteristics of the response group are presented in Table 1.

7.3.1 Validity

Ten percent of the children scored above the adapted cut-off point. Eight percent had a parent-identified – definite or severe – difficulties score on the impact supplement of the SDQ and six percent had a high score for emotional symptoms, conduct problems or hyperactivity in combination with a high impairment score. Thirteen percent was classified by any of the three other classification methods. Thirteen percent were classified as having problems identified by any of the three other classification methods. The Kappa coefficient measuring the agreement between these three scoring methods varied from 0.49 to 0.59, which means a moderated agreement (0 is no agreement and 1 is perfect agreement). Four percent of the children had an elevated score on all three scoring methods.

Of all children 8% had a clinical CBCL TPS, 10% had a clinical CBCL Internalizing score, 6% a clinical Externalizing score, and 7% were being treated for psychosocial problems. Table 2 presents the AUC, sensitivity and specificity indices for each of the four classification methods. AUC values ranged from 0.68 to 0.86 for the CBCL criteria and from 0.69 to 0.75 for the treatment status as criterion. For each criterion, the 95% confidence intervals overlap. So, the overall performance of the four classification methods does not differ significantly. The highest sensitivity for the identification of a clinical CBCL TPS was found for the combination score (0.80). The combination score was also the most sensitive for a clinical CBCL Internalizing score and Externalizing score. However, this score had the lowest specificity (varying from 0.90 to 0.92).

Almost all 95% confidence intervals for sensitivity overlapped, meaning that these differences in sensitivity are not significant. We found two significant differences: compared to parent-defined problems, the combination method is more sensitive to a clinical CBCL TPS; compared to high-subscale/impairment it is more sensitive to a clinical Internalizing Problems score.

Table 1   *Social and demographic indicators and the percentage of children with problems indicated by four classification methods: Elevated Total SDQ Problems (SDQ TDS[#]), Parent-reported difficulties(PR-Dif), High subscale / impairment (Sub-Imp) and a Combination of these three (Comb)***

	%	SDQ TDS %	PR-Dif %	Sub-Imp %	Comb %
Total (n=707)		9.8	8.2	6.4	13.6
Gender child		ns	*	ns	**
Boy	49	11.9	10.7	7.5	17.7
Girl	51	7.7	5.8	5.2	9.7
Age of the child		ns	ns	ns	ns
7 years	11	4.0	1.3	1.3	5.3
8 years	7	11.3	15.1	5.7	20.8
9 years	28	10.6	9.6	7.1	14.1
10 years	22	10.1	8.2	5.7	14.5
11 years	23	10.6	6.2	7.5	12.4
12 years	9	9.8	11.5	9.8	16.4
Ethnic background		ns	ns	ns	ns
The Netherlands or other industrialized country	83	9.2	8.0	6.6	13.2
Non-industrialized country	7	15.2	10.9	6.5	17.4
Unknown	10	11.1	8.3	4.2	13.9
Family situation		ns	ns	ns	
Two parents	86	9.1	8.1	5.6	13.0
One parent	10	16.4	9.6	12.3	19.2
Unknown	3	11.1	11.1	11.1	16.7
Parenting situation		ns	*	*	ns
2 biological parents	81	8.7	7.1	5.2	12.2
1 biological parent	14	16.2	13.1	11.1	21.2
Other/unknown	5	8.8	11.8	11.8	14.7
Work situation		*	ns	**	ns
No job	3	13.6	13.6	13.6	13.6
One fulltime job	30	8.0	9.0	3.8	12.3
One part-time job	4	7.4	0.0	3.7	7.4
One fulltime + one part time job	42	9.1	7.0	6.4	12.8
Both part time	8	5.0	10.0	3.3	13.3
Both fulltime	3	11.1	5.6	0.0	11.1
Unknown	10	21.4	11.4	17.1	24.3
Education		*	ns	ns	*
(None) elementary school	3	21.1	15.8	0.0	31.6
Lower education	25	12.9	8.4	7.3	16.9
Medium education	32	10.1	8.8	5.7	13.6
Higher education	33	5.5	6.8	5.9	9.3
Unknown	6	13.0	8.7	10.9	15.2

#: using an adapted TDS cut-off point, ≥ 14
*: Chi-Square: p<0.5; **= p<0.01; ***=p<0.001

Table 2   Test characteristics of the SDQ, using clinical CBCL scores and treatment status as criterion; Area under Curve (AUC), Sensitivity, Specificity and 95% Confidence Intervals (95% CI), in relation to four classification methods (n=707).

	AUC (95% CI)	Sens. (95% CI)	Spec. (95% CI)
CBCL Total Problems			
SDQ Total Difficulties score[a, b]	0.85 (0.79-0.92)	0.75 (0.63-0.86)	0.96 (0.94-0.97)
Parent-reported difficulties[c]	0.73 (0.64-0.81)	0.50 (0.37-0.63)	0.95 (0.94-0.97)
High sub-scale / impairment [d]	0.75 (0.66-0.83)	0.52 (0.39-0.65)	0.97 (0.96-0.99)
Combination of three scores[e]	0.86 (0.80-0.93)	0.80 (0.70-0.91)	0.92 (0.90-0.94)
CBCL Internalizing Problems			
SDQ Total Difficulties score[a, b]	0.73 (0.66-0.81)	0.51 (0.40-0.63)	0.95 (0.93-0.97)
Parent-reported difficulties[c]	0.71 (0.63-0.79)	0.46 (0.34-0.57)	0.96 (0.95-0.98)
High sub-scale / impairment [d]	0.68 (0.60-0.76)	0.39 (0.28-0.50)	0.97 (0.96-0.99)
Combination of three scores[e]	0.79 (0.72-0.86)	0.65 (0.54-0.76)	0.92 (0.90-0.94)
CBCL Externalizing Problems			
SDQ Total Difficulties score[a, b]	0.83 (0.75-0.92)	0.73 0.59-0.86)	0.94 (0.92-0.96)
Parent-reported difficulties[c]	0.72 (0.62-0.82)	0.50 (0.35-0.65)	0.94 (0.93-0.96)
High sub-scale / impairment [d]	0.75 (0.65-0.84)	0.53 (0.37-0.68)	0.96 (0.95-0.98)
Combination of three scores[e]	0.83 (0.75-0.91)	0.75 (0.62-0.88)	0.90 (0.88-0.92)
Treatment Status			
SDQ Total Difficulties score[a, b]	0.69 (0.60-0.79)	0.46 (0.32-0.60)	0.93(0.91-0.95)
Parent-reported difficulties[c]	0.70 (0.61-0.79)	0,46 (0.32-0.60)	0.94 (0.93-0.96)
High sub-scale / impairment [d]	0.66 (0.56-0.75)	0.35 (0.22-0.49)	0.96 (0.94-0.97)
Combination of three scores[e]	0.75 (0.67-0.84)	0.60 (0.46-0.74)	0.90 (0.87-0.92)

a Using an adapted Clinical SDQ cut-off; ≥ 14   d Percentage above cut-off: 6.4%
b Percentage above cut-off: 9.8%   e Percentage above cut-off: 13.6%
c Percentage above cut-off: 8.2%

7.3.2 Added value

Table 3 presents odds ratios (OR) which express the likelihood of a clinical score for the CBCL scales or for 'currently being treated' if a child was classified as having a problem by one of the four classification methods, after taking the identification of problems by the health professional into account. Adding any of the SDQ based classification methods into the equation always led to a significant change in the log likelihood ratio. The ORs for all the classification methods were significant, regardless of the criterion used. Overall, the SDQ improves the identification of children with an elevated CBCL Internalizing Problems score less well than the identification of children with an elevated clinical CBCL Total Problems and Externalizing Problems score. An elevated SDQ TDS had most added value for the prediction of a clinical CBCL compared to the other three classification methods. Parent-reported difficulties added most to the prediction of 'currently being treated'. However, once again, the 95% Confidence Intervals for al criteria overlapped.

7.4 Discussion and conclusion

This study assessed the suitability of the SDQ for the early detection of psychosocial problems among children aged 7-12 years by the PCH. We looked at the validity of four SDQ-based classification methods: 1) the SDQ TDS, 2) the (definite or severe) difficulties reported by the parents using the impact supplement of the SDQ, 3) an

elevated score for emotional symptoms, conduct problems, hyperactivity in combination with an elevated impairment score and 4) a combination method: an elevated score for any of these three classifications. The results show that all four scoring methods of the SDQ are valid and have added value for the identification of psychosocial problems among children. We found that the SDQ TDS and the combination method (which includes the elevated TPS) were most sensitive for elevated CBCL scores and that the parent-identified difficulties and the combination method were most sensitive for children currently being treated. However, most differences between the scoring methods, were not statistically significant. The exception was the combination method, which was significantly more sensitive for an elevated CBCL TPS than the scoring method based on parent-reported difficulties. The combination method was also statistically more sensitive to an elevated CBCL Internalizing Problems Score compared to an elevated score for emotional symptoms, conduct problems or hyperactivity in combination with a high impairment score.

Finally, the SDQ TDS added most to the identification of psychosocial problems by the PCH, although the differences between the classification methods were again not statistically significant.

Table 3   Results from multiple logistic regression analyses of the four classification methods on clinical CBCL scores and treatment status, taking the identification by the PCH into account (n=707);adjusted odds ratios (OR) and 95% Confidence Intervals (95% CI), for identification by PCH and the SDQ classification methods

SDQ classification method	PCH detects problems (yes vs. no) OR (95% CI)	Elevated SDQ classification method* OR (95% CI)
Clinical CBCL Total Problems score		
Total Difficulties Score yes (vs. no) *	2.59 (1.23-5.47)	49.93 (23.67-105.36)
Parent-reported Difficulties yes (vs. no)	3.57 (1.85-6.89)	16.33 (6.20-24.45)
Subscales/ Impairment score yes (vs. no)	3.47 (1.76-6.85)	26.76 (12.53-57.17)
Combination score yes (vs. no)	1.96 (0.95-4.06) ns	36.04 (16.66-77.96)
Clinical CBCL Internalizing Problems score		
Total Difficulties Score yes (vs. no) *	3.58 (2.01-6.40)	12.92 (6.97-23.92)
Parent-reported Difficulties yes (vs. no)	3.56 (1.99-6.35)	12.52 (6.52-24.03)
Subscales/ Impairment score yes (vs. no)	3.96 (1.80-6.29)	13.79 (6.74-28.21)
Combination score yes (vs. no)	2.63 (1.43-4.85)	15.17 (8.26-27.88)
Clinical CBCL Externalizing Problems score		
Total Difficulties Score yes (vs. no) *	1.92 (0.85-4.32) ns	31.69 (13.96-71.98)
Parent-reported Difficulties yes (vs. no)	2.79 (1.30-5.97)	10.44 (4.85-22.48)
Subscales/ Impairment score yes (vs. no)	2.55 (1.17-5.57)	19.64 (8.81-43.81)
Combination score yes (vs. no)	1.73 (0.78-3.83) ns	21.29 (9.24-49.06)
Being treated for psychosocial Problems		
Total Difficulties Score yes (vs. no) *	7.20 (3.46-14.99)	5.54 (2.78-11.06)
Parent-reported Difficulties yes (vs. no)	6.72 (3.20-14.11)	7.02 (3.44-14.34)
Subscales/ Impairment score yes (vs. no)	7.72 (3.72-16.00)	5.68 (2.66-12.11)
Combination score yes (vs. no)	5.75 (2.70-12.25)	6.75 (3.40-13.41)
* Cut-off point : equal to or higher than 14		

Bourdon et al.[17] found significant differences between the classification methods: an elevated SDQ TDS alone distinguished less well between children with service contact/use than parent-reported difficulties and an elevated score for emotional symptoms, conduct problems, or hyperactivity-inattention in combination with an elevated impairment score. By contrast, we found no significant differences in sensitivity or added value between the SDQ TDS and the other classification methods. This may be due to the fact that our study sample was much smaller. The number of cases in our sample was therefore rather small and the power of tests for sensitivity and for the OR in the logistic regressions was therefore rather small. However, tests for differences between AUC indices are much less sensitive to small numbers and we found no differences here, either.

The percentage of children scoring above the UK cut-off point in our study was lower than in the UK. In Germany and the United States the 10% cut-off point ($\geq$ 16) also tended to be slightly lower than in the UK, but it was much closer to the UK than the cut-off point in this study.[17,14,20] Another Dutch study, however, found the same 10% cut-off point of 14 for the parent SDQ as this study did.[21] The authors of the other Dutch study concluded that a possible reason for this lower cut-off was the substantial level of non-response among parents (response was 63%).[15,21] In the present study the response was much higher (87%) and the effect of non-response is therefore smaller. A study of the CBCL showed that Dutch parents also reported fewer problems on the CBCL than US parents, but this did not apply to German parents, suggesting that it is a structural pattern.[22] We therefore believe that the lower SDQ scores in the Netherlands are not the result of some flaw in the study, but that they reflect a higher level of well-being among Dutch children (compare, for example, a recent report from UNICEF[23]).

The SDQ scores in this general population sample are most sensitive for a CBCL Total Problem Score and least sensitive to internalizing problems and current treatment. The impact supplement enhanced the identification of internalizing problems slightly, but sensitivity remained lower than for the total problems score. This concurs with the findings of Goodman, who indicates that 'Not surprisingly, the algorithm seems most likely to miss children with relatively encapsulated symptoms that are not well covered by the SDQ'. It is important to mention that Goodman refers to a multi-informant algorithm (parent, teachers, and self-reports from older children) in which he found a greater likelihood of missing encapsulated or internalising problems. He proposes that "if researchers or clinicians want to detect as many emotional or hyperactivity disorders as possible, they would be well advised to use the SDQ prediction for "any disorder" rather than for "emotional disorders" or "hyperactivity disorders". A second-stage screening procedure can then be used to detect which SDQ 'positive' children have the disorder of particular interest" (Goodman et al.[24], pages 537 and 538).

Strengths and limitations

This study has important strengths but also some limitations. Its strengths are the national coverage and high response rates. One limitation is that the evaluation of questionnaires for emotional and conduct problems is always hampered by difficulties in the choice of a gold standard: there is simply no definitive indicator of such problems. This study therefore adopted a common strategy to overcome this problem: in our study we included both the CBCL and current treatment as validation criteria. One of the

problems, however, is that both the CBCL and the SDQ are completed by parents. This probably leads to higher correlation between these two instruments because both instruments rely on the opinion of the parents. Clinical assessments, like psychiatric interviews, do not suffer from these problems and could therefore be more convincing as a criterion. However, we could not use psychiatric interviews in this study, because of the costs and burden for the parents. Skovgaard et al.[25] also indicate that screening of a whole population can be conducted using an instrument such as the CBCL and that diagnostic classification should take place in a second stage with a combination of psychometric and clinical approaches. These clinical assessments are expensive and time-consuming and should be restricted to smaller samples, consisting of, for example, individuals identified by screening procedures, such as the CBCL. In the Netherlands, PCH is an important service for the identification of these high-risk children in the population as a whole. The CBCL is technically adequate for this first step in the identification of psychosocial problems, but it is too long, too time consuming and therefore too costly and not suitable for use in PCH. The extent to which another shorter instrument can replicate the global classification of the CBCL is then a valid measure of the suitability of this instrument.

At the same time, the inclusion of both treatment status and the CBCL as criterion measures in our study is, in our view, a major advantage, compared to the study of Bourdon et al., since using contact with services as the only criterion neglects the fact that many children with serious problems never contact services because of their problems.

Policy Implications

The results of this study show that use of the SDQ can provide effective support for the PCH in the identification of psychosocial problems among children. The routine use of an instrument of this kind in PCH is therefore recommended. For a first identification of children with problems by the Dutch PCH, the use of the SDQ Total Difficulties score is justified, since more complicated and time consuming scoring and classification methods do not significantly improve identification.

What this paper adds

The SDQ as a short instrument for the detection of psychosocial problems among children can provide effective support for the identification of these problems in preventive child healthcare. For a first identification, the use of only the SDQ Total Difficulties Score is justified.

References

1   Verhulst FC, Ende van der J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997;36:901-9.

2   Geeraert L, Noortgate W van den, Grietens H, Onghena P. The effects of early prevention programs for families with young children at risk for physical child abuse and neglect: a meta-analysis. *Child Maltreat* 2004;9(3):277-91.

3   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by Preventive Child Healthcare. *Arch Pediatr Adolesc Med* 2001;155: 462-9.

4   Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems among toddlers in Dutch preventive Child Healthcare. *Arch Pediatr Adolesc Med* 2004;158 (august): 811-17.

5   Zeijl E, Crone M, Wiefferink K, Keuzenkamp S, Reijneveld M. *Children in the Netherlands*. [in Dutch] SCP/TNO, The Hague, The Netherlands, 2005.

6   Vogels T, Reijneveld SA, Brugman E, den Hollander-Gijsman ME, Verhulst FC, Verloove-Vanhorick SP. Detecting psychosocial problems among 5-6 years old children in Preventive Child Healthcare: The validity of a short questionnaire used in an assessment procedure for detecting psychosocial problems among children. *Eur J Public Health* 2003;13: 353-360.

7   Reijneveld SA, Vogels AGC, Brugman E, Van Ede J, Verhulst FC, Verloove-Vanhorick SP. Early detection of psychosocial problems in adolescents: how useful is the Dutch short indicative questionnaire (KIVPA)? *Eur J Public Health* 2003; 13: 152-159.

8   Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry* 1997; 38:581-6.

9   Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. *J Child Psychol Psychiat* 1999; 40: 791-9.

10   Hawes DJ, Dadds MR. Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. Australian and New Zealand *J Psychiat* 2004;38: 644-651.

11   Malmberg M, Rydell AM, Smedje H. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nord J Psychiatry.* 2003;57(5):357-63.

12   Smedje H, Broman JE, Hetta J, von Knorring AL. Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *Eur Child Adolesc Psychiatry.* 1999 Jun;8(2):63-70.

13   Koskelainen M, Sourander A, Kaljonen A. The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *Eur Child Adolesc Psychiatry.* 2000 Dec;9(4):277-84.

14   Woerner W, Becker A, Rothenberger A. Normative data and scale properties of the German parent SDQ. *European Child Adolescent Psychiatry* 2004; 13 (suppl 2); II/3-II/10.

15   Widenfelt BM van, Goedhart AW, Treffers PDA, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003;12(6):281-9.

16   Muris P, Meesters C, van den Berg F. The strengths and difficulties questionnaire (SDQ). Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child & Adolescent Psychiatry* 2003;12:1-8.

17   Bourdon K.H., Goodman R, Rae DS, Simpson G, Koretz DS. The strengths and difficulties questionnaire: U.S. normative data and psychometric properties. J Am *Acad Child Adolesc Psychiatry*. 2005 Jun;44(6):557-64.

18   Achenbach TM. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. University of Vermont Department of Psychiatry, Burlington, 1991.

19   Verhulst FC, van der Ende J, Koot HM. *Handleiding voor de CBCL/4-18 [Manual for the CBCL/4-18]*. Erasmus University/Department of Child and Adolescent Psychiatry, Sophia Children's Hospital, Rotterdam, The Netherlands, 1996.

20   Mojtabai R. Serious emotional and behavioral problems and mental health contacts in American and British children and adolescents. *J Am Acad Child Adolesc Psychiatry*. 2006 Oct;45(10):1215-23.

21   Goedhart A, Treffers F, van Widenfelt B. Asking about psychosocial problems among children and adolescents. The Strengths and Difficulties Questionnaire (SDQ).[in Dutch] *Maandblad Geestelijke volksgezondheid* 2003; 58:1018-35.

22   Crijnen AA, Achenbach TM, Verhulst FC. Problems reported by parents of children in multiple cultures: the Child Behavior Checklist syndrome constructs. *Am J Psychiatry* 1999 Apr;156(4):569-574.

23   UNICEF. *Child poverty in perspective: an overview of child well-being in rich countries*. Innocenti Report Card 7. 1-52. 2007. Florence, Unicef Innocenti Research Center.

24   Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiat* 2000; 177:534-9.

25   Skovgaard AM, Houmann T, Landorph SL, Christiansen E. Assessment and classification of psychopathology in epidemiological research of children 0-3 years of age. A review of the literature. *Eur child Adolesc Psychiatry* 2004; 12: 337-46.

8 Item Response Theory based Computerized Adaptive Testing can provide an accurate and efficient identification of children with psychosocial problems

A.G.C. Vogels

G.W. Jacobusse

S.A. Reijneveld

Abstract

Objectives

To assess if a Computerized Adaptive Test can overcome the weaknesses of short written questionnaires when identifying children with psychosocial problems.

Research design

We used an existing national data set with 205 items on behavioral and emotional problems (n=2041). In a first random subsample we determined which items met the requirements of an IRT model sufficiently. Using those items, item parameters necessary for a Computerized Adaptive Test were calculated and a cut-off point was defined. In the second subsample we determined the validity and efficiency of a Computerized Adaptive Test in a simulation study, with current treatment status and the Total Problem Scale of the Child Behavior Checklist as criteria.

Subjects

Respondents were 2041 parents of children invited for a routine health examination by Preventive Child Healthcare (PCH) (response: 84%).

Results

Out of 205 items available only 15 did not sufficiently meet the criteria of the underlying IRT model. For 90% of the children a score above or below cut-off point could be determined with 95% accuracy. Sensitivity and specificity with the Total Problem Scale as a criterion were 0.89 and 0.91. The mean number of items needed to achieve this was 12.

Conclusion

An IRT-based CAT is a very promising option for the identification of psychosocial problems in children, as it can lead to an efficient, yet high-quality identification. The results of our simulation study need to be replicated in a real-life administration of our Computerized Adaptive Test.

8.1 Introduction

Many children suffer from behavioral and emotional problems,[1-3] and these problems may seriously interfere with their daily functioning, now and later in life.[4,5] Yet many of these children remain untreated.[6] Durlak and Wells showed that early identification considerably improves the prognosis of the children involved.[7]

Community-based preventive child healthcare (PCH) services, especially those working outreachingly, are in a unique position to identify such problems as early as possible. In the Netherlands, PCH professionals offer routine well-child care to the entire Dutch population to the age of about 14. The early detection of children with psychosocial problems is an explicit part of their working package. In contrast to systems existing e.g. in the US, Dutch PCH does not offer treatment services. When (physical or psychosocial) problems are detected, children are referred to other parts of the healthcare system, especially to primary healthcare. Research has shown, however, that early identification in PCH is often far from perfect. For example, Brugman et al. showed that in Dutch PCH, about half of the children with a clinical CBCL Total Problem Score remained unnoticed when they were examined by a physician or nurse.[1] Other studies came to similar conclusions.[8-11]

There are several possibilities to improve the identification of children with emotional and behavioral problems. Wiefferink et al. showed that using clear protocols and extensive staff training can lead to a significant increase in the number of children with problems identified and a decrease in the number of children incorrectly identified as having problems.[12] Other studies showed that using good questionnaires, to be filled in by parents, teachers or the children themselves, can also help to improve the quality of early identification.[7,13,14] However, in community-based PCH , the time available for each individual child is limited. This means that questionnaires that are practicable in such settings, have to be easy to score and therefore short. Also, they must be easy for all parents to answer. Short questionnaires, unless they have a very narrow scope, tend to be less reliable and less valid than desirable.[15] Identification of problems based on such questionnaires is therefore error prone, resulting in too many false classifications.

Since the 1950s, a series of new statistical models called Rasch or IRT (Item Response Theory) models have been developed which allow for Computerized Adaptive Testing (CAT), a short and efficient test procedure that does not compromise the accuracy of the test results. Originally these models could only be applied to items with only two categories. This limited their application mainly to the field of intelligence testing and the assessment of school achievements.[16] In the last decades more widely applicable models became available. This led to IRT-based test procedures in the field of quality of life measurements.[17] Very recently some publications have been published describing the application of these models to the assessment of mental health problems.[18-20]

In this study we assessed whether CAT can also be used for a fast, short, yet high-quality identification of children with emotional and behavioral problems in community-based PCH.

8.2 Methods

8.2.1 General introduction

Just like test procedures based on more traditional psychometric theories, IRT-based procedures help to determine the position of a person on some measurement scale, for instance on intelligence, school achievements or the level of psychosocial problems. In IRT that position is called the person location. IRT differs from traditional psychometrics in that it provides information about which items are relevant to use in an individual assessment and which are less useful.

A simple example may illustrate this principle: suppose in a particular arithmetic test, a child failed to give the correct answer to the question: how much is 2*3. In that case it is probably not very useful to ask: how much is 34*17. The latter question can help to distinguish between children on a higher position of the arithmetic ability scale, but will add little information for a child who failed the first question. Translating this to scales assessing emotional en behavioral problems, items indicating severe problems are not informative for children with no or few problems and items indicating less severe problems are not informative for children with severe problems.

With IRT it is possible to determine the severity of individual items; *i.e.* the position on the scale where it is informative. That position is called the item location.[16] This information can be used to shorten the test length in the following way. After each answer on a single question an estimation is made of the person's probable score, or person location. Then the available items are scanned in order to determine which item could improve the estimated person location. This continues until a previously defined accuracy has been reached. In practice this process is only possible with the aid of computers: Computerized Adaptive Testing (CAT).[21] For CAT to be possible, the location of the items must be known in advance, before actual testing of an individual starts.

In order to assess whether such a procedure is viable and efficient for the identification of children with problems, three questions need to be answered:

1. Are the items of four questionnaires on emotional and behavioral problems suitable for an IRT-based CAT and, if so, what are parameters of the individual items, to be used in a CAT?

2. Which cut-off point results in a sensitive and specific distinction?

3. What are the validity and the specificity of such a CAT and how efficient is this procedure?

8.2.2 Data collection, population and measures

We used a data set collected in an earlier study[22] containing information about parent-reported problems of children aged seven to twelve. Data were collected in a two-step procedure. In the first step nine regional PCH organizations were found willing to participate in our study. Second, parents who were invited for a routine health examination of their child were asked to participate in the study and to fill in some questionnaires about emotional and behavioral disorders of their child.

Data from 2041 parents were available, 84% of the parents invited to participate. 51% of the children involved were girls. Mean age was 10.1 (sd=1.4). In 83% of all cases both parents were born in the Netherlands and in 6% at least one parent came from a non-OECD country or from Turkey. Fourteen percent did not live in a two-parent family and only three percent lived in a family were none of the parents was employed. The sample was representative for the population of this age group under care in Dutch PCH, regarding gender, age, ethnic origin and family composition (two parents, one parent, other), with Cohen's W varying from .002 for gender to .109 for ethnic origin.

Each parent answered the Child Behavior Checklist (CBCL) and one out of three questionnaires: the Pediatric Symptom Checklist (PSC),[23-25] the Strengths and Difficulties Questionnaire (SDQ),[26-28] or a newly developed Dutch questionnaire on psychosocial problems for children in primary education, the PSYBOBA.[29] More information on the relatively unknown PSYBOBA may be obtained from the authors. The PSC, SDQ and PSYBOBA were chosen for this study because there was evidence for their conceptual validity in relation to the kind of problems Dutch PCH aims to identify and because they met the requirements for use in the context of PCH: short, easy to administer and to score. The data collection led to an incomplete data matrix: the data for the PSC, the SDQ and the PSYBOBA are each available in about one third of the sample.

Finally, data on current treatment status for emotional or behavioral problems were obtained from the PCH professionals. These answered questions regarding this, based on medical records and on the routine health examination of the child, during which a small structured interview was done for the purpose of this study. One of the questions asked whether the child was currently being treated for mental health problems and, if so, by which kind of professional or institution.

8.2.3 Analyses

We randomly divided the total sample in two subsamples. The first one, the calibration sample (n=1650), was used to answer the first two questions (suitability of the items and determination of the cut-off point). The second, the validation sample (n=391), was used for the evaluation of the validity and efficiency. This evaluation in a separate sample was done in order to prevent overestimation of validity and efficiency coefficients.

In order to assess the suitability of the items for an IRT-based CAT we assessed whether the items fitted the assumption of unidimensionality. For this aim, we determined whether the items showed enough fit with the Partial Credit Model (PCM), one of the unidimensional IRT models. Using this model for a CAT has the advantage that it results in scores on an interval measurement level.[30] We performed this assessment using the RUMM 2020 software (http://www.rummlab.com.au/)[31], as this can handle incomplete data matrices like ours. RUMM 2020 provides so called outfit statistics for each item, that indicate to what extent each item fits the model. Items were considered suitable for CAT measurement if they had an outfit statistic smaller than 1.7.

Next, we calculated the item locations of the remaining items, using the same software. In order to determine whether the estimated item locations would be valid, independently from gender and ethnicity, we performed Differential Item Functioning

(DIF) analyses for each item. We did this by multinomial logistic regressions, with the raw score on the item as the dependent variable. First, the estimated person location was the only predictor in the logistic regression model. Second, both gender, ethnicity and their interaction were added as predictors. Items were considered as showing DIF when these additional predictors had a significant effect and led to an increase of the explained variance of more than 3.5%.[32]

Second, we determined an optimal cut-off point for the CAT scores, i.e. one which enables a good distinction between a non-clinical versus a clinical CBCL TPS. The CBCL TPS was used as the criterion measure, because it measures exactly the emotional and behavioral problems which Dutch PCH aims to identify and because both its concurrent and predictive validity have been widely established.[33-37] We simulated a CAT in the calibration sample, using the answers on paper and pencil questionnaires as if they were given in a CAT and calculated the resulting person locations (CAT scores). We assume that in community-based PCH about 30 items is the maximum number feasible, and limited the number of items to be used in this CAT to 30. We used Fisher's information Index[38] for the selection of the next item in the CAT. Using the scores from this simulation we did a Receiver Operating Characteristics analysis with a clinical CBCL TPS as criterion and chose that point that resulted in a specificity of 0.90 as cut-off point.

The exact estimate of the person locations, however, will vary somewhat with the number of items used in the CAT. In order to assess the effect of this variation we repeated the analyses with a fixed number of 5, 10 and 20 items and also with no limit to the number of items, but continuing until the person locations had been estimated with 95% accuracy. In all these CATs the first item was chosen at random. We calculated the sensitivities and specificities for all these analyses and inspected the differences, in order to verify that the maximum of 30 items we used was a sensible one.

Finally, we evaluated the validity and efficiency of the CAT. The validity was assessed by means of a simulated CAT in the independent validation sample. In this simulation we aimed to assess, with an accuracy of 95%, whether a person scored above or below the chosen cut-off point. In other words, the CAT was stopped when the 95% Confidence Interval of the estimated person location did not overlap anymore with the chosen cut-off point. This procedure is known as clinical decision adaptive testing.[39] Again, the starting item was chosen at random and Fisher's Information index was used to select the next best item. We assessed the validity of the estimated person locations by calculating the Area Under Curve (AUC), sensitivity and specificity with a clinical CBCL TPS and current treatment status as criteria.

The efficiency of the procedure was evaluated by calculating the number of items needed in this simulated CAT and the number of respondents for whom the CAT resulted in 95% certainty on a score below or above the chosen cut-off point.

For the IRT analyses we used the RUMM 2020 software[31]. All other analyses were done using SPSS.

8.3 Results

8.3.1 Are the items on emotional and behavioral problems suitable for an IRT-based CAT?

Of the 205 items in the four questionnaires 190 met the criteria for a CAT: they had an outfit of less than 1.7. Most items that had to be removed came from the CBCL (13 out of 15). The Person Separation Index was 0.93, indicating a high reliability. The DIF analyses showed that almost all estimates were not modified by gender and ethnicity. Only 8 of the 190 items showed some DIF; five items in relation to gender (sexual problems, running away, attacking others, being ill without physical cause and problems with teachers) and three in relation to ethnicity (tantrums, not being assertive, talking about suicide). Most of these problems have a very low prevalence and will therefore have a small overall impact on the final estimations. We therefore decided not to remove these items.

Figure 1 presents the estimated item locations calculated for the remaining items and split by questionnaire. As mentioned before, these item locations are indications of the level of severity. The most severe items on the right (concerning very serious problems) were items from the CBCL, which in general appeared to have more severe items than the other three questionnaires.

Figure 1   Estimated locations of the items in the four questionnaires in the calibration sample

117
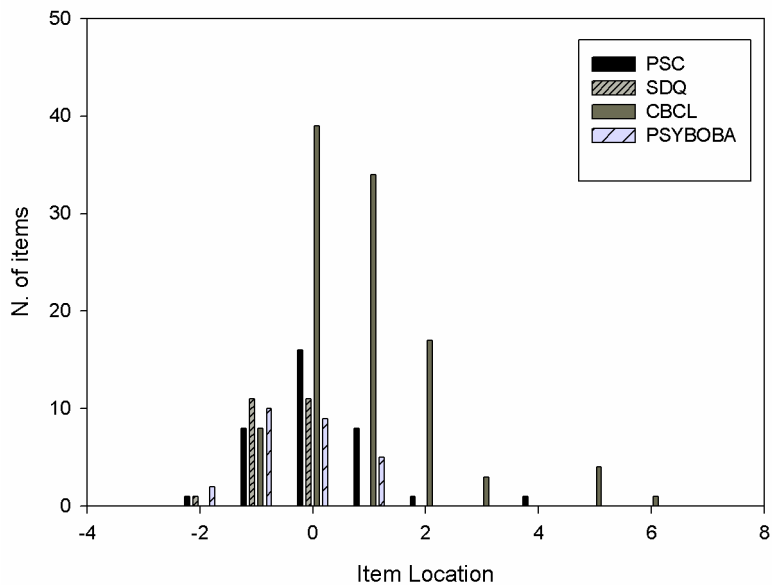
8.3.2 Determining the cut-off point

After the item locations had been estimated, we did a CAT simulation on the calibration sample with a fixed number of 30 items. Figure 2 presents the number of respondents by the calculated person location on the latent scale, by CBCL TPS, divided into normal, borderline or clinical. The ROC analysis showed that with a cut-off point of -1.9 the specified specificity of 0.90 was reached. The sensitivity for a clinical CBCL TPS at that point was 93%.

Table 1 presents the effects in terms of AUC, sensitivity and the specificity indices in relation to the use of different numbers of items in the CAT. The specificity shows little variation; using a fixed number of 5 or 10 items results in a decreased sensitivity. The results for a CAT with 20 or 30 items and for a CAT that continues until the 95% Confidence Interval does not overlap with the cut-off point, are very similar.

Figure 2   *Distribution of estimated person locations in the calibration sample by CBCL classification*
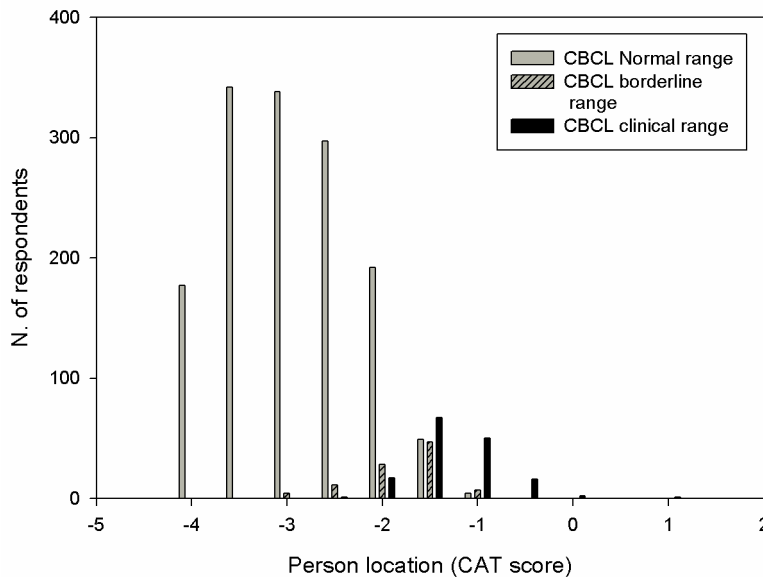

Table 1   *Sensitivity and specificity at a cut-off point of -1.9 in relation to the maximum number of items used to estimate persons' locations.*

Criteria to stop the CAT	Sensitivity	Specificity	Area Under Curve
Max. no of items: 30	93%	90%	0.97 (0.96 – 0.98)
Max. no of items: 20	90%	89%	0.96 (0.95 – 0.97)
Max. no of items: 10	79%	88%	0.92 (0.90 – 0.94)
Max. no of items: 5	64%	88%	0.90 (0.84 – 0.89)
Estimation of person location with 95% accuracy	92%	90%	0.97 (0.96 – 0.98)

118

8.3.3 Validity and efficiency

In the validation sample, the ROC analyses showed that the CAT did very well in the identification of children with a clinical TPS; the AUC was 0.92 (CI: $0.85 - 0.99$). With the chosen cut-off point sensitivity was 0.89 (CI: $0.71 - 0.97$), with a specificity of 0.91 (CI: $0.87 - 0.93$). Kappa was 0.53.

Using treatment status as criterion the AUC was 0.74 (CI: $0.63 - 0.84$). The sensitivity for current treatment status was 0.55 (CI: $0.37 - 0.72$), with a specificity of 0.89 (CI: $0.85 - 0.92$). Kappa was 0.32.

Overall, in relation to the CBCL TPS, the CAT selection procedure resulted in a correct classification of 91% of all children involved. The CAT resulted in a correct classification for the large majority of cases with normal (96%) or clinical scores (89%). However, 20 (77%) of the 26 cases with a score in the CBCL borderline range, had an elevated CAT score.

Figure 3 presents the number of items needed to reach convergence, i.e. to assess with 95% certainty whether the respondents had a true score below or above the chosen cut-off point of -1.9. In 40 cases (10%) convergence was not possible with less than 100 items. They had a mean person location of -1.88 (sd=0.18); i.e. very near the chosen cut-off point. Their mean CBCL TPS was 28.4 (sd=7.1).; 25% of them had a CBCL TPS in the borderline range; 5 % in the clinical range.

For the other 351 cases, the mean number of items used was 11.5 (sd=13.0). For 37% of the respondents the procedure converged with less than 5 items; for 57% up to 9 items were needed. For 74% up to 20 items were used and for 82% up to 30 items.

Figure 3   Efficiency of the CAT procedure: percentage of persons for whom a score above or below the cut-off point could be estimated with 95% accuracy, by number of items used to achieve convergence.
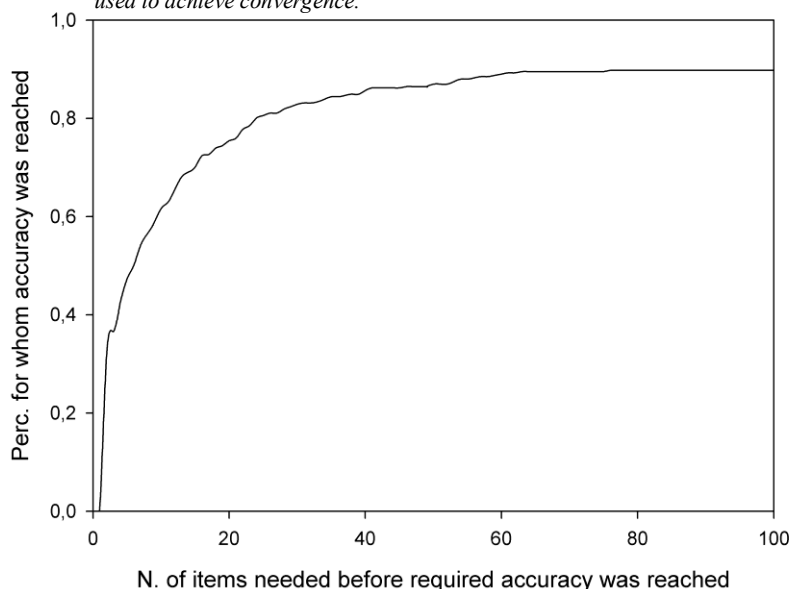
The mean CBCL TPS for respondents for whom less than 5 items were used in the CAT was 10.8 (sd=10.5). We checked the convergence between the CBCL TPS based classification and the CAT classification for these respondents. In 98% of the cases the classification was identical. The CAT resulted in a below cut-off point score for 2 respondents with a clinical CBCL TPS and one respondent got a CAT score above the cut-off point with a CBCL in the normal range.

8.4 Conclusion

This study showed that IRT-based Computerized Adaptive Testing indeed resulted in an accurate, yet very efficient identification of children with psychosocial problems. Most of the items of the four questionnaires investigated met the requirements of an IRT model, needed to incorporate them in a CAT. A simulation study showed that the procedure identified children with a clinical CBCL TPS with high sensitivity and specificity. For 90% of all cases we could determine with 95% certainty whether they had an elevated score. In order to achieve these results, on average only 11.5 items were needed. For more than half of the children less than ten items were needed.

There are, of course, other, more traditional techniques for reducing test length. However, other than more traditional approaches, an IRT-based CAT provides high measurement quality, by adjusting items used in the assessment to the individual being tested. This has the additional advantage that this individual is not being confronted with items that are not relevant in his situation and that possibly may be shocking for him or her. The inclusion of items of the SDQ, PSC and SDQ in the item pool used for this CAT offers therefore the advantage that more items are available that are suitable for parents of children with no or few problems.

Fit with the literature

Our finding that an IRT-based CAT can result in accurate assessments with far less items than tests based on traditional psychometrics is fully comparable to findings in other studies, applying IRT CAT techniques in the fields of intelligence and school achievement assessment,[15,21] and in the field of Quality of Life.[17,40]

The first studies on the application of IRT models in the field of the identification of psychosocial problems have now been published,[18-20] and these studies came to similar conclusions. Compared to other studies, known to us and regarding CAT and mental health, our study and the study by Gardner[19] are the only ones to focus on a rather broad concept, rather than on more specific problems, like Gardner[18] and Fliege et al.[20] Gardner[19] used the PSC as criterion. As we used the more widely validated CBCL as one criterion, our study is a stronger argument for the usefulness validity of CAT-based procedures in the field of mental health.

Gardner[19] evaluated the extent to which a multidimensional adaptive test could be used to replicate screening decisions based on the Pediatric Symptom Checklist. He found a very high correspondence between the Adaptive PSC and the original 35 items PSC (kappa=0.84), higher than the corresponding figure we found. The mean number of items he needed to achieve this was 12 items, out of 35, whereas we needed a mean of

12 items, to replicate the screening decision based on the 120 item CBCL. It is not exactly clear why he found a higher correspondence than we did. Our cut-off point was not chosen in order to maximize kappa, but had we done so, our kappa would still be lower than Gardner's. One explanation may be that Gardner used a multidimensional model, whereas we used the simple one-dimensional Partial Credit Model. Another explanation might be that Gardner limited himself to PSC items, whereas we used items from four questionnaires. Thus, in our study there is less overlap between the items in the CAT and the criterion measure. This is probably the main reason that Gardner's study resulted in a higher kappa.

Strengths and Limitations

This study has several strengths but also limitations. A major strength is that it concerns a community-based sample of children with high response rates and representative for the population under care. Moreover, the pool of items used was much broader than the criterion that we used, leading to a relative independence of the criterion from the items used for prediction. This was reinforced by the use of separate samples for the construction of the CAT and for its validation. However, our study also has several limitations. Though more independent than in previous studies, some of the predicting items will also be included in the criterion. This is a limited problem, though, if one tries to reach the best short alternative for a longer questionnaire, as we did. Moreover, we simulated a CAT based on answers given to a full questionnaire, which is a deviation from the real practice set-up. A next stage will certainly be to evaluate the CAT in a real-life situation. Finally, although we had a rather large sample, our validation sample was relatively small, implying the need for a large scale replication. Anyhow, our study provides a valid assessment of the potentials of an IRT-based CAT for PCH practice.

Implications

The most important implication of our study is that IRT-based CAT appears to be a very feasible and promising tool to improve the identification of psychosocial problems in PCH. As such it earns a quick passing through to the daily practice of well-child care and maybe even of pediatric care in general. Before having a final pass to clinical practice, several aspects have to be studied more thoroughly, though. This in particular concerns the use of our simulated version in a real-life situation, with parents filling out the CAT on real computers. Currently, a beta-version for this aim is available at www.uwkind.nl, but this is Dutch only and protected by passwords and firewalls to preserve patient-confidentiality. A formal assessment of this implementation in daily practice is the next step for research. Similarly, our findings have to be replicated in other settings and maybe using other item pools as well. Anyhow, this new technology may provide a push to improve the quality of the identification of psychosocial problems in PCH.

References

1   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001 April;155(4):462-9.

2   Kelleher KJ, McInerny TK, Gardner WP, Childs GE, Wasserman RC. Increasing identification of psychosocial problems: 1979-1996. *Pediatrics* 2000 June;105(6):1313-21.

3   Horowitz LM, Wang PS, Koocher GP et al. Detecting suicide risk in a pediatric emergency department: development of a brief screening tool. *Pediatrics* 2001 May;107(5):1133-7.

4   Moffitt TE, Caspi A. Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Dev Psychopathol* 2001;(13):355-75.

5   Verhulst FC, Van der Ende J., Rietbergen A. Ten-year time trends of psychopathology in Dutch children and adolescents: no evidence for strong trends. *Acta Psychiatr Scand* 1997 July;96(1):7-13.

6   Verhulst FC, Ende van der J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997 July;36(7): 901-9.

7   Durlak JA, Wells AM. Primary prevention mental health programs for children and adolescents: a meta-analytic review. *Am J Community Psychol* 1997 April;25(2):115-52.

8   Glascoe FP. Using parents' concerns to detect and address developmental and behavioral problems. *J Soc Pediatr Nurs* 1999 January;4(1):24-35.

9   Lavigne JV, Binns HJ, Christoffel KK et al. Behavioral and emotional problems among preschool children in pediatric primary care: prevalence and pediatricians' recognition. Pediatric Practice Research Group. *Pediatrics* 1993 March;91(3):649-55.

10   Costello EJ, Janizewski S. Who gets treated? Factors associated with referral in children with psychiatric disorders. *Acta Psychiatr Scand* 1990;81:523-9.

11   Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems among toddlers in Dutch preventive Child Healthcare. *Arch Pediatr Adolesc Med* 2004;158(august 2004):811-7.

12   Wiefferink CH, Reijneveld SA, de Wijs J, Swagerman M, Campman D, Paulussen TG. Screening for psychosocial problems in 5-6-year olds: a randomised controlled trial of routine health assessments. *Patient Educ Couns* 2006 January;60(1):57-65.

13   Reijneveld SA, Vogels AG, Brugman E, van Ede J, Verhulst FC, Verloove-Vanhorick SP. Early detection of psychosocial problems in adolescents: how useful is the Dutch short indicative questionnaire (KIVPA)? *Eur J Public Health* 2003 June;13(2):152-9.

14   Vogels T, Reijneveld SA, Brugman E, Hollander-Gijsman M, Verhulst FC, Verloove-Vanhorick SP. Detecting psychosocial problems among 5-6-year-old children in preventive Child Healthcare: the validity of a short questionnaire used in an assessment procedure for detecting psychosocial problems among children. *Eur J Public Health* 2003 December;13(4):353-60.

15   Embretson SE, Reise SP. *Item Response Theory for psychologists*. London: Lawrence Elbaum Associate Publishers; 20060.

16   Wright D, Masters G. *Rating Scale Analysis*. Chicago Il: Mesa Press; 1982.

17   Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 1997 August;6(6):595-600.

18   Gardner W, Shear K, Kelleher KJ et al. Computerized adaptive measurement of depression: a simulation study. *BMC Psychiatry* 2004 May 6;4:13.:13.

19   Gardner W, Kelleher KJ, Pajer KA. Multidimensional adaptive testing for mental health problems in primary care. *Med Care* 2002 September;40(9):812-23.

20   Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res* 2005 December;14(10):2277-91.

21   Wainer H. *Computerized Adaptive Testing. A Primer*. Hillsdale: Lawrence Erlbaum Associates, Publishers; 1990.

22   Vogels AGC, Crone MR, Hoekstra F., Reijneveld SA. Drie vragenlijsten voor het opsporen van psychosociale problemen bij kinderen van zeven tot twaalf jaar. Leiden: TNO Kwaliteit van leven; 2005.

23   Jellinek M, Evans N, Knight RB. Use of a behavior checklist on a pediatric inpatient unit. *J Pediatr* 1979 January;94(1):156-8.

24   Jellinek M, Little M, Murphy JM, Pagano M. The Pediatric Symptom Checklist: support for a role in a managed care environment. *Arch Pediatr Adolesc Med* 1995 July;149(7):740-6.

25   Jellinek M, Murphy JM, Robinson J, Feins A, Lamb S, Fenton T. Pediatric Symptom Checklist: screening schoolage children for psychosocial dysfunction. *J Pediatr* 1988;112(2):201-9.

26   Bourdon KH, Goodman R, Rae DS, Simpson G, Koretz DS. The Strengths and Difficulties Questionnaire: U.S. normative data and psychometric properties. *J Am Acad Child Adolesc Psychiatry* 2005 June;44(6):557-64.

27   Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Int Rev Psychiatry* 2003 February;15(1-2):166-72.

28   Goodman R, Scott S. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *J Abnorm Child Psychol* 1999;27(1):17-24.

29   Ede van J, Rover Cd, Verhappen MJ, Vries-Lequin Id. *PSYBOBA. A pilot study on the usability of a questionnaire for psychosocial problems among primary school children aged 7 till 12 [in Dutch]*. GGD Nederland; 2003.

30   Andrach D. *Rasch models for measurement*. Newbury Park: Sage Publications; 1988.

31   Andric D, Luo G. Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. J Appl Meas 2003; 4(3): 205-221. 2003.

32   Jodoin MG, Gierl MJ. Evaluating Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education* 2001;14:329-49.

33   Koot HM, Oord van den EJ, Verhulst FC, Boomsma DI. Behavioral and emotional problems in young preschoolers: cross-cultural testing of the validity of the Child Behavior Checklist/2-3. *J Abnorm Child Psychol* 1997 June;25(3):183-96.

34   Koot HM, Verhulst FC. Prediction of children's referral to mental health and special education services from earlier adjustment. *J Child Psychol Psychiatry* 1992 May;33(4):717-29.

35   Verhulst FC, Dekker MC, Van der EJ. Parent, teacher and self-reports as predictors of signs of disturbance in adolescents: whose information carries the most weight? *Acta Psychiatr Scand* 1997 July;96(1):75-81.

36   Verhulst FC, Van der EJ. Six-year stability of parent-reported problem behavior in an epidemiological sample. *J Abnorm Child Psychol* 1992 December;20(6):595-610.

37   Verhulst FC, Koot HM, Enden van der J. Differential predictive value of parents' and teachers' reports of children's problem behaviors: a longitudinal study. *J Abnorm Child Psychol* 1994 October;22(5):531-46.

38   Vestering N, Buuren Sv. *Two New Bayesian item selection methods and their application in Computerized Classification Testing*. Leiden: TNO Quality of Life; 2004.

39   Waller NG, Reise SP, Computerized adaptive personality assessment: an illustration with the Absorption scale. *J Pers Soc Psychol* 2006;57(6):1051-8.

40   Ware JE Jr, Bjorner JB, Kosinski M. Practical implications of item response theory and computerized adaptive testing: a brief summary of ongoing studies of widely used headache impact scales. *Med Care* 2000 September;38(9 Suppl):II73-II82.

9 The identification by Dutch Preventive Healthcare of children with psychosocial problems: do short questionnaires help? Discussion and implications

A.G.C. Vogels

9.1 Main findings

In Chapter 1 the main questions we sought to answer in this thesis were formulated as:

1. Can differences in the number of children identified as having psychosocial problems by individual health professionals be explained by differences in the prevalence of problems or background characteristics in the groups of children they examine?

2. What are the psychometric qualities of a number of short questionnaires PCH could use to identify children with psychosocial problems and could they improve the identification by PCH?

3. Is it possible to develop a Computerized Adaptive Test, using items from questionnaires on psychosocial problems, in order to achieve a short, yet accurate assessment of the likelihood of psychosocial problems being present?

These questions can now be answered.

Ad 1: Differences between individual health professionals

In Chapter 2 we showed that there are significant differences between individual PCH professionals in the number of children they identify as having problems. Moreover, it was shown that these differences cannot be explained by differences in the prevalence of problems nor by the background characteristics of the children they examined.

Ad 2: Can short paper-and-pencil questionnaires improve the identification of children with problems by PCH?

In Chapters 3 to 7 we evaluated a number of short paper-and-pencil questionnaires and tried to determine whether such questionnaires could help PCH to improve the identification of children with problems. We evaluated five questionnaires given to three different age groups:

- The LSPPK (National Checklist for Indicating Psychosocial Problems in Five Year Olds, Landelijke Signaleringslijst voor Psychosociale Problemen bij Kleuters), for parents of preschool children

- The PSYBOBA (Questionnaire for psychosocial problems for primary-school children, Vragenlijst voor PSYchosociale problematiek in de BOvenbouw van het BAsisonderwijs), the Pediatric Symptom Checklist (PSC) and the Strengths and Difficulties Questionnaire (SDQ) for parents of children aged 7 to 12

- The KIVPA (Short Indicative Questionnaire for Psychosocial Problems among Adolescents, Korte Indicatieve Vragenlijst voor psychosociale problematiek bij adolescenten).

We showed that all these questionnaires could indeed improve the identification by Dutch PCH. However, the KIVPA and LSPPK versions which were evaluated do so less well than the other questionnaires in the 7 to 12 age group.

Ad 3: Is a CAT an option?

In Chapter 8 we showed that most of the items from the PSYBOBA, the PSC, the SDQ and the Child Behavior Checklist (CBCL) could be used for an Item Response Theory based Computer Adaptive Test. In a simulation study we also demonstrated that such a

CAT resulted in sensitive and specific distinction between children with or without problems. As expected, this distinction could be made very efficiently, using on average less than 12 items for each individual assessment.

9.2 Discussion

Differences between individual PCH professionals

We showed that there are systematic differences between individual PCH professionals regarding the proportion of children they identify as having problems. These differences could not be explained by differences in problems or risk indicators among the children seen by PCH professionals. We know of no other studies that have used a similar methodology to assess differences between clinical judges. Therefore we cannot compare our results with those of other studies, in the field of PCH or in other settings. It would be interesting to investigate this issue in settings like community-based pediatric centers and primary healthcare, in which staff members with limited psychological and psychiatric training have to identify children or adults with problems. It seems likely that such studies would result in similar conclusions.

This does not mean, however, that our results should not be taken seriously. We showed that differences in clinical judgment about the presence of psychosocial problems are greatest for children with a CBCL Total Problem Score (TPS) at or somewhat above the clinical cut-off point of the CBCL. For children with a TPS of 40 the model predicted probability of being identified as having a problem is on average about 40%. For half of the professionals this probability rate varies between 30% and 50%, an for the other half the variation is even greater.

The identification of children with problems is one of the aims of the uniform part of PCH's Basic Working Package. This program should be offered to all children in a standardized way. This means, among other things, that the chance of problems being detected should be, as far as possible and feasible, equal and not depend on the particular professional examining the child. Earlier studies[1,2] showed that PCH did not identify problems in a large percentage of children for whom CBCL data indicated a great likelihood of problems being present when they attempted to identify those children unaided by short questionnaires. Moreover, they identified problems in many children who according to the CBCL were unlikely to have problems. This in itself is no proof that the identification of problems by Dutch PCH must be wrong. The CBCL indicates the likelihood of the presence of problems and the individual health professional may be right in arriving at a different judgment than the CBCL data would suggest.

However, if this were the main reason for the discrepancies between CBCL results and the conclusions of PCH professionals, one would expect such discrepancies to be randomly distributed between individual PCH professionals. We showed that this was not the case. We showed that the chance of a child with a given problem level being identified as having a problem varies, depending on which PCH health professional is examining this particular child.

Clearly, this means that the identification of children with problems based on clinical judgment during the routine health examination, without the use of validated instruments to support this identification, lacks the degree of standardization required for the Uniform Part of PCH's Basic Working Package.

Evaluating a single questionnaire or comparing available questionnaires

In chapter 6 we reported a randomized comparison between three questionnaires that seemed suitable for use by PCH for children aged 7 to 12. Despite the fact that the results showed that the psychometric qualities of these three questionnaires were quite comparable and that they all could help to improve the identification of psychosocial problems, we feel that comparing available questionnaires, whenever possible, is preferable to an evaluation of a single questionnaire. We know of no other studies that used a similar design. Our approach is comparable to what is now rapidly becoming standard in studies assessing effectiveness and economic evaluations of interventions. Such studies do not try to assess the effectiveness or costs as such, but compare specific interventions with other interventions or with usual care. Such an approach is far more helpful in guiding health policy decisions, such as deciding which instruments to use. We feel that this type of a comparative approach is also worthwhile in the evaluation of questionnaires to be used in other areas of healthcare. Only a systematic comparison can guarantee that the best instrument available will indeed be chosen.

Treatment status and CBCL or YSR as criteria to assess the validity of the questionnaires evaluated in this thesis

All studies presented in this thesis, except the study in chapter 2, assessed the validity of short instruments when used as a method to identify children with psychosocial problems. In chapter 1 psychosocial problems were defined indicatively as emotional and behavioral problems. For the evaluation of instruments that should help improve the identification of such problems we used current treatment status, referral and high scores on the CBCL or the Youth Self Report (YSR) as criteria.

Current treatment status (or more specifically, having been treated during the last 6 months by some professional or institution because of social and behavioral problems) is a criterion which very much resembles criteria used in validation studies: being referred to some form or care. Herjanic and Campbell[3] for example used being referred to psychiatric care as validation for the Diagnostic Interview for Children and Adolescents (DICA), a standardized interview developed for the assessment of psychopathology in children, both for clinical and epidemiological use. Yet, criteria such as referral and treatment status should be handled with some care. Research has shown that many children with serious problems are not treated for them.[4,5] This may be caused by all kind of factors, such as lack of available facilities, problems not being identified as such, parents who are able to cope with problems without feeling the need for professional help and so on. Also, not all of the children who are being treated will have very serious problems. Anxious parents may seek help for their children far sooner than less anxious parents. It should also be remembered that family doctors or PCH professionals may refer children with less competent parents more often than they would in the case of children with highly competent parents.

Therefore, one cannot expect very high indices of association between current treatment status or referral and instruments identifying problems in children. Even a perfect instrument would show rather imperfect validity, when validated against referral or treatment status. In other words: instruments aiming at identifying children with problems must by expected to show some relationship to treatment status, but one cannot expect very high associations to be found. This, in general, is exactly what we found for the instruments evaluated in this study.

The second criterion we used was the CBCL or the YSR – for adolescents. This is an accepted approach in the literature[6] and is especially recommended for large-scale studies.[7] The reason we chose these measures as a criterion was twofold. First, numerous studies in many different countries and cultures have shown that these instruments are highly valid. They are strong indicators for concurrent problems and are very predictive of future problems and negative developmental outcomes. Few other rating instruments have been so widely validated, if any at all. The second reason that we chose these instruments is that there are few viable alternatives for rating scales. One might consider using a full-blown psychiatric interview or some form of structured psychiatric interview. Such methods are very time-consuming and expensive and were therefore not feasible for these studies, neither is their validity perfect. Edelbrock et al.,[8] for example, found that the one-week test-retest reliability of the Diagnostic Interview Schedule for Children (DISC) for children aged 6 to 9 years old was only 0.39. He observed a sharp decline in the number of symptoms reported on the second occasion. Boyle et al.[9] evaluated the Diagnostic Interview for Children and Adolescents for use in general population samples and concluded, among other things, that interview data provided by 6 to 11 year olds to classify internalizing disorders were too unreliable to be useful and that agreement between parent-child/adolescent dyads was generally low. Granero Perez et al.[10] did research in order to explain the low test-retest reliability of the same instrument and found both characteristics of the child and characteristics of the questions to be relevant here. Jensen[11] found lower test-retest kappa coefficients in a community sample than in a clinical sample.

In discussions with representatives from the PCH field, the use of the CBCL or YSR as criterion measures was sometimes criticized because these measures should be seen as focused on psychiatric disorders and PCH must identify not only children with specific psychiatric disorders, but all children with emotional and behavioral problems in need of support. These questionnaires were indeed developed as measures to be used in the context of child psychiatry. However, a psychiatric disorder is diagnosed when a defined combination of symptoms is present, in combination with burden for the patient. Questionnaires like the CBCL and YSR do not assess psychiatric disorders in this sense. They collect information on a variety of aspects of a child's functioning and all these aspects must be considered as relevant when the phrase 'psychosocial problems' is used as indicated by the expert meeting on PCH and Psychosocial Problems in 1999 and Blokland et al.[12] School problems or cognitive problems are also included in their definitions of psychosocial problems. The CBCL and YSR only assess attention problems, ignoring other cognitive problems. In this respect the CBCL and YSR may be less adequate criteria.

This thesis shows that elevated scores on short questionnaires are clearly associated with high scores on the CBCL or the YSR. This, then, justifies the expectation that elevated scores on the questionnaires evaluated will be associated with concurrent and future problems, too. Yet, as the CBCL is not a perfect criterion, but only an indicator of a strong likelihood of the presence of problems, data from instruments validated with the CBCL must be interpreted with care. Suppose, an elevated SDQ score detects 79% of cases with a borderline CBCL TPS[13] and a borderline CBCL TPS has a sensitivity of 63% of all children referred to mental health services.[14] Then – other things being equal – the sensitivity of the SDQ for being referred for treatment must be expected to be still lower. Similarly, if the specificity of a borderline CBCL TPS is 0.84 and the specificity of the SDQ for a borderline TPS is 0.90, then the specificity of the SDQ – other things being equal – must also be expected to be lower. This means, that a relative large proportion of cases – defined in terms of being referred to mental health institutions – may remain undetected with the SDQ. Similarly, the proportion of non-cases with a false positive elevated SDQ score will also be considerable.

These figures underline the necessity of a very careful interpretation of data from the questionnaires evaluated in this study. They also underline, that these instruments should be used as a first tool, to be used in the context of the standardized health examination and supplemented, as far as possible, by other validated assessment methods and not as a pre-selection tool, used to select those children who need to be seen by PCH.

Further evaluations of the questionnaires validated in this study, using other criteria, need to be carried out. PCH centers are in a unique position to perform such studies, as they use these instruments among large numbers of children and PCH is able to collect information on concurrent and predictive validity. The introduction of the Electronic Health Dossier can facilitate such studies and may contribute to a higher level of a scientifically founded PCH.

Generalizability of the results

All studies presented in this thesis, except the study in Chapter 2, assessed the validity and added value of short instruments when used as a method to identify children with psychosocial problems. The validity of measuring instruments is not an inherent characteristic of the instrument as such, but is related to the population among which and the situation in which it will be used. The studies presented were all done among largely representative samples of the Dutch population examined by PCH. Moreover, data collection took place in the normal, daily practice of PCH, before and during the normal routine health examinations, and thus reflected the situation in which the instruments will be used. This means that the conclusions of these studies may be generalized to the actual practice in Dutch PCH in which these instruments will be used. However, a word of warning is in order here because there are three problems that must be born in mind.

The samples were largely representative. However, the percentage of parents and children from ethnic minority groups participating in these studies is smaller than in the population as a whole. This is a problem, since there are indications that ethnicity is a factor related to the problems reported by parents and to the identification by Dutch

PCH. Reijneveld et al.,[15] for example showed that the prevalence of parent-reported problems on the CBCL is higher among children from the former Dutch colonies and children from economic immigrants than among children from the indigenous population. They also showed that PCH identified more problems among children whose parents were economic immigrants and that there was no relation between identification and the CBCL Total Problem Score for this group, whereas there was a clear association among indigenous children. Therefore, the under-representation of children from ethnic minorities may mean that the generalizability cannot be extended to include this entire group.

The under-representation of children/parents from migrant communities may have been caused by three factors. The first is that parents and children from ethnic minority groups may show up less often when invited for a routine PCH health examination. In as far as this is the case, it does not affect the representativeness of the data for the population that does receive PCH care. A second factor may be that parents from ethnic minority groups do show up for routine health examinations, but are not willing or not able to answer paper-and-pencil questionnaires. Vogels et al.[13] showed that 16% of parents from non-OECD countries who showed up for the routine health examination, did not fill in the questionnaires; for parents of Dutch origin this was only 8%. Clearly, our findings are not shown to be valid for children and parents not participating for this reason. Moreover, although a large majority of non-OECD parents did participate, our findings may not yet be considered as valid for these parents either: the number of these parents participating was far too small to determine with any acceptable degree of confidence the sensitivity of the questionnaires for these groups. The third factor that may have contributed to the under-representation of ethnic minorities in our studies is that PCH services in the largest cities did not participate in our studies. Clearly, this also limits the generalizibility of our findings.

A second word of warning concerns the fact that the studies in this thesis evaluated the instruments as questionnaires to be answered prior to a routine health examination. This reflects the way in which most PCH departments use such questionnaires. Sometimes, however, they are used as a pre-selection tool in which only parents or children with a score above a certain cut-off point are invited for a further examination. It may very well be that parents answer the questionnaires in a situation like that with a different attitude from when they know that a standard examination will follow. Based on this thesis, using the questionnaires evaluated prior and in preparation for a routine health examination may be considered as evidence-based; using them as a pre-selection tool may not.

The third warning to bear in mind concerns the way the questionnaires are actually being used as part of the standard health examination. The instruments were evaluated as a method to identify children who may have problems, using certain cut-off points. An elevated score on these questionnaires can be compared to a signal like an alarm clock going off and we determined the sensitivity and specificity of that signal as an indicator for possible problems. Discussions and conversations with individual PCH professionals, however, strongly show a considerable variation in the way in which these questionnaires are used in current day-to-day practice. Sometimes, the cut-off point is changed, sometimes it is not used at all, for example because the individual

professional simply does not know that there is a defined cut-off point. Some professionals tend to consider some items as far more important than other. Sometimes the phrasing of questions or answering categories is changed, based on the feeling that the new phrasing is better. Once again, using adapted questionnaires or using them in a way different from the way they have been evaluated here, must be considered as non evidence-based. It will probably lead to more individual variation between professionals and to a lack of standardization for this part of the PCH's Basic Working Package.

So, some restrictions can and should be made regarding the generalizability of the results of the studies presented in this thesis, especially regarding children from ethnic minorities and regarding possible deviations in day-to-day practice from the way these instruments should be used. Despite these restrictions, our results can be considered valid for the very large majority of the population under PCH care in the age of 5 to 14. Overall, the proper use of these instruments will enhance the identification of children with problems.

The validity and feasibility of a CAT need to be determined in a real-life situation

Chapter 8 explored the possibilities of an IRT-based Computer Adaptive Test for the identification of children with problems. The results strongly suggested that such a procedure could result in a very accurate identification, similar to or better than short paper-and-pencil questionnaires evaluated in this thesis. We also showed that such a procedure is very efficient, needing, on average less than twelve questions to determine whether a child scores above or below a chosen cut-off point. As a consequence, most parents need to spend only a few minutes filling them in before this result is achieved. This presents the opportunity of asking them for additional information that can be rated on more specific scales, such as Internalizing and Externalizing problems.

Our findings were based on a simulation study. Therefore, the results of this study have not yet proven that such a CAT would be a valid and efficient procedure in day-to-day practice as well. Therefore, before this CAT is implemented, there must be a trial run to evaluate the procedure in real life.

An IRT-based CAT can only be used if a computer is available. This may be a problem, as at the moment, not all parents can use computers. Furthermore, computerized tests offer not only more accurate and efficient measurement; they can also automatically calculate scale scores, so the PCH professional does not need to spend time doing the scoring himself.

In the meantime an Internet-based application has been developed. This will be used in a study that will determine the validity and efficiency in real life and will explore which conditions must be met in order to be able to implement this procedure effectively.

The implementation of questionnaires in day-to-day practice needs attention

Overall, we may conclude that the questionnaires evaluated in this thesis can help PCH to improve the identification of children with psychosocial problems. What we did not determine is whether they actually do so, when used in daily PCH practice. This is not self-evident. Earlier we mentioned that the way in which questionnaires are used by different organizations and different individual PCH professionals varies. This probably reduces the actual added value these questionnaires are capable of providing when used for the identification of children with problems. The fact that the Dutch association of

regional health centers (GGD Nederland) has decided to develop a systematic procedure for the implementation of the SDQ in PCH for 7- to 12-year olds is an important step forward and should be extended to other age groups and other questionnaires. Further research is needed to assess exactly how questionnaires are being used in practice in relation to the question whether their use actually improves the identification of children in need of help. Only then will we be able to label this part of the Uniform Basic Working Package as evidence-based.

9.3 Implications

Implications regarding PCH practice

Clearly, validated questionnaires can improve the identification of children with psychosocial problems. So, PCH should use them in daily practice, whenever possible and feasible. It should ensure that available questionnaires are used as intended. That means that the questionnaires evaluated in this thesis should be used in combination with a high-quality assessment during the routine health examination and not as a pre-selection tool. It also means that such questionnaires should not be changed, even if there are arguments to do so, for example because the phrasing could be improved here and there. Even small changes can seriously impair the validity of a questionnaire.[16] It also means that cut-off points, defined on the basis of research evidence, need to be taken seriously and adhered to.

PCH professionals who are going to use questionnaires have to be taught how to use them, how to interpret the results and what to do when they find an elevated score. This needs to be monitored by some form of quality management. GGD-Nederland has to be complimented for its efforts to develop a well-considered program for the implementation of the SDQ in PCH practice. Such programs should be expanded to encompass other questionnaires, too.

This study addressed the identification of children with problems in the context of standard health examinations. This is an essential part of PCH's Basic Working Package. However, the number of routine health examinations for children aged 4 and older is limited to just a few and could be considered insufficient for tracking down problems occurring between two consecutive health examinations. Currently, Care and Advice Teams (Zorg- en AdviesTeams, ZAT) are being set up in and around schools for primary and secondary education. These ZATs are an important safety-net for children, providing a service allowing for a more or less continuous identification of children with problems. PCH is expected to participate in these ZATs but its participation is not considered as being part of the uniform part of the Basic Working Package. This anomaly should be addressed as soon as possible.

Unfortunately, when a child is identified as having a problem and in need of help, such help is not always available, or not at the appropriate time. Referring to the criteria developed by Wilson and Junger[17] this fact is sometimes used as an argument against the need to develop or to use good identification methods in PCH. This is both short-sighted and in defiance of one of the essential responsibilities of PCH in the

Netherlands. It is short-sighted because high-quality identification not only means that children with problems will be referred, but also that fewer children without serious problems will be referred. A high-quality identification therefore helps to ensure that the available capacity in mental healthcare and other services is used for those children who need it most.

The argument that high-quality identification procedures are less urgent because health services do not have the capacity needed to handle these problems is also in defiance of one of PCH's essential responsibilities: PCH is obliged to monitor Public Health, not only for the immediate benefit of individual children, but also in order to provide authorities with information about what kind of health problems need attention and which kind of care has to be developed to ensure that those in need of help can indeed receive such help. Only high-quality identification procedures can result in reliable and valid prevalence data which can be used by policy makers.

Based on this thesis and the discussion of the findings the following recommendations can then be made:

1. For the identification of children with problems PCH should not rely on the clinical judgment of individual professionals alone, but should use methods and instruments which have been shown to be valid.

2. It should use these instruments as they were validated and for the purpose for which they were evaluated; changes should only be allowed after new validation studies.

3. The introduction of instruments in PCH practice should be carefully planned and should be embedded in a continuous implementation and quality management program, ensuring that all staff members know how to use them and continue to use them as intended.

4. A lack of adequate youth healthcare services should be considered as a major argument in favor of a high-quality methods to identify children with problems, not used as an argument against it.

5. The identification of children with problems in the ZAT-context must be seen as an essential addition to their identification in the routine health examination. The participation of PCH in these ZATs must be redefined and included in the Uniform Part of the Basic Working Package.

Implications for research

This thesis reported on the validity and added value of several questionnaires for the identification of children with psychosocial problems in PCH services provided for children aged 5 and 6 , 7 to 12 and adolescents. Unfortunately, for children under 5 no validated questionnaires are available. Data clearly indicate that although PCH for younger children identifies many children as having problems, it also misses many children with serious indications that problems are present. PCH is aware of this problem and repeatedly expressed a need for such questionnaires. Therefore, research into ways to filling this gap is clearly needed. Also, although studies showed that the KIVPA and the LSPPK questionnaires can improve the identification, they were shown to have serious shortcomings. In the 7 to 12 age group the SDQ performed better.

Research to assess whether it would be feasible to replace the LSPPK and the KIVPA by the SDQ would therefore be desirable, the more so as this would lead to comparable data between the different assessment moments.

Dutch PCH has a legal obligation to monitor the health of all children from 0 to 19. However, for adolescents above 14 years of age no systematic monitoring or care is provided. This gap needs serious reconsideration. Methods need to be developed and tested to ensure that youngsters in this age group who are in need of support are actually identified as well. Internet-based techniques, based on Computerized Adaptive Tests like the one described in Chapter 7, may be very helpful for this purpose.

It is now generally agreed that using information on psychosocial problems from one source only will inevitably lead to an incomplete picture.[18,19] Most studies reported in this thesis, however, relied solely on one source, either the parent or the child. The reason for this was that until now, a standardized collection of data from more than one source has not been common practice in PCH. At the moment several PCH centers are considering including multiple-source data collection as part of their care for individual children. At least one center does so already. Internet-based computerized adaptive tests may prove an attractive method in realizing such a multiple-source data collection, as they require a minimal amount of effort from respondents (for example, teachers).

As mentioned earlier, this thesis showed that short questionnaires may improve the identification of children with problems by PCH, but did not prove that they actually do so when used in daily practice. Available evidence suggests that the way in which questionnaires are used varies considerably between individual professionals and PCH centers. This means that more attention should be paid to the actual implementation of questionnaires in daily practice and whether they are used in such a way that they will effectively improve the identification of children with psychosocial problems.

Elevated scores on short questionnaires like those evaluated in this thesis are valid indicators of the likelihood of the presence of certain problems. However, they offer no absolute certainty, neither do they provide enough information to assess the exact nature of the problem, its seriousness and the need for professional support. For this to be achieved a further assessment is required. At least part of this assessment, inevitably, has to be performed by PCH itself, as it would not be feasible to refer all children with an elevated score to a Youth Care Office (Bureau Jeugdzorg, BJZ). Until now, no standardized methods have been developed for PCH to perform such assessments. This may be a contributory factor in variation between individual PCH professionals. Thus, there is a clear need for the development and evaluation of assessment methods assessment to be used by PCH when short questionnaires indicate the likelihood of the presence of problems. Methods which may be considered here are, for example, hiring personnel with specific expertise, using some form of standardized interviews aiming at the assessment of problems, or other assessment systems, like the Development And Well-being Assessment procedure.[20]

The identification of psychosocial problems by PCH in a well connected chain of care with all its partners is one of the current political priorities. The first identification of children with psychosocial problems, the assessment needed after the first signal, and the determination of the kind of help or support needed, all cost time. So does co-

operation between institutions. How much time is not very clear, and it is necessary to determine what extra time is needed for children who are identified as having problems. Then, if the political priorities are to be taken seriously, funding for such a Working Package should be made available.

So the following recommendations are made:

1. Methods and instruments need to be developed for all groups for which such instruments are not yet available and/or validated, especially children under 5 and adolescents over 14.

2. The quality of the identification of children with problems among ethnic minorities needs to be assessed.

3. Instruments and methods need to be developed and validated which allow for a systematic assessment by PCH of the nature and seriousness of problems and the need for professional help, when short questionnaires indicate the likelihood of problems.

4. Methods need to be developed to collect data from more than one informant; especially since the Internet and computer adaptive testing may be promising methods in this respect.

5. The question to what extent and under which conditions short questionnaires do actually improve the identification of and care for children with problems in daily practice needs to be addressed.

6. The validity and feasibility of a Computer Adaptive Test in day-to-day PCH practice needs to be investigated.

7. Evaluating the relative merits of a number of comparable questionnaires is to be preferred over the evaluation of a single questionnaire.

References

1   Brugman E, Reijneveld SA, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems by preventive child healthcare. *Arch Pediatr Adolesc Med* 2001 April;155(4):462-9.

2   Reijneveld SA, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Identification and management of psychosocial problems among toddlers in Dutch preventive Child Healthcare. *Arch Pediatr Adolesc Med* 2004;158(august 2004):811-7.

3   Herjanic B, Campbell W. Differentiating psychiatrically disturbed children on the basis of a structured interview. *J Abnorm Child Psychol* 1977;5(2):127-34.

4   Verhulst FC, Ende van der J. Factors associated with child mental health service use in the community. *J Am Acad Child Adolesc Psychiatry* 1997 July;36(7):901-9.

5   Verhulst FC, Berden GF, Sanders-Woudstra JA. Mental health in Dutch children: (II). The prevalence of psychiatric disorder and relationship between measures. *Acta Psychiatr Scand Suppl* 1985;324:1-45.:1-45.

6   Widenfelt van BM, Goedhart AW, Treffers PD, Goodman R. Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 2003 December;12(6):281-9.

7   Skovgaard AM, Houmann T, Landorph SL, Christiansen E. Assessment and classification of psychopathology in epidemiological research of children 0-3 years of age: a review of the literature. *Eur Child Adolesc Psychiatry* 2004 December;13(6):337-46.

8   Edelbrock C, Costello AJ, Dulcan MK, Kalas R, Conover NC. Age differences in the reliability of the psychiatric interview of the child. *Child Dev* 1985 February;56(1):265-75.

9   Boyle MH, Offord DR, Racine Y et al. Evaluation of the Diagnostic Interview for Children and Adolescents for use in general population samples. *J Abnorm Child Psychol* 1993 December;21(6):663-81.

10   Granero PR, Ezpeleta AL, Domenech Massons JM, de la Osa CN. Characteristics of the subject and interview influencing the test-retest reliability of the Diagnostic Interview for Children and Adolescents-Revised. *J Child Psychol Psychiatry* 1998 October;39(7):963-72.

11   Jensen P, Roper M, Fisher P et al. Test-retest reliability of the Diagnostic Interview Schedule for Children (DISC 2.1). Parent, child, and combined algorithms. *Arch Gen Psychiatry* 1995 January;52(1):61-71.

12   Blokland G. Groepsgerichte opvoedingsondersteuning. *Nederlands tijdschrift voor jeugdzorg* 2002;(2):97-103.

13   Vogels AGC, Crone MR, Hoekstra F., Reijneveld SA. Drie vragenlijsten voor het opsporen van psychosociale problemen bij kinderen van zeven tot twaalf jaar. Leiden: TNO Kwaliteit van leven; 2005.

14   Verhulst FC, Koot HM. *Manual for the CBCL/4-18 [in Dutch]*. Rotterdam: EUR/Sophia Kinderziekenhuis; 1996.

15   Reijneveld SA, Harland P, Brugman E, Verhulst FC, Verloove-Vanhorick SP. Psychosocial problems among immigrant and non-immigrant children--ethnicity plays a role in their occurrence and identification. *Eur Child Adolesc Psychiatry* 2005 May;14(3):145-52.

16   Goodman R, Iervolino AC, Collishaw S, Pickles A, Maughan B. Seemingly minor changes to a questionnaire can make a big difference to mean scores: a cautionary tale. *Soc Psychiatry Psychiatr Epidemiol* 2007 April;42(4):322-7.

17   Wilson JMG, Jungner G. Principles an Practice of screening for disease. Geneva: World Health Organization; 1968.

18   Verhulst FC, Dekker MC, Van der EJ. Parent, teacher and self-reports as predictors of signs of disturbance in adolescents: whose information carries the most weight? *Acta Psychiatr Scand* 1997 July;96(1):75-81.

19   Goodman R, Ford T, Corbin T, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) multi-informant algorithm to screen looked-after children for psychiatric disorders. *Eur Child Adolesc Psychiatry* 2004;13 Suppl 2:II25-II31.

20   Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000 July;41(5):645-55.

10 Summary

This thesis focuses on the early detection by Dutch Preventive Child Healthcare (PCH) of children with psychosocial problems and especially on the question whether short questionnaires can improve this early detection.

The following questions were answered:

1. Can differences in the number of children identified as having psychosocial problems by individual health professionals be explained by differences in the prevalence of problems or background characteristics in the groups of children they examine?

2. What are the psychometric qualities of a number of short questionnaires PCH could use to identify children with psychosocial problems and could they improve the identification by PCH?

3. Is it possible to develop a Computerized Adaptive Test, using items from questionnaires on psychosocial problems, in order to achieve a short, yet accurate assessment of the likelihood of psychosocial problems being present?

In accordance with the conclusion of an expert meeting in 1999 on the prevention of psychosocial problems among children and adolescents by PCH, the term psychosocial problems will be used in this thesis to indicate emotional and behavior problems and problems at school.

Introduction

Chapter 1 describes how PCH started to focus on psychosocial problems more explicitly in the seventies of the last century. In the beginning the extent to which and the way how this was done was largely dependent on local interests, leading to large differences in identification and management. This changed under the influence of the National Working Group on the Early Detection of Psychosocial Problems (Landelijke Werkgroep Signalering van Psychosociale Problematiek bij Jongeren, LSPPJ) and the introduction of the PCH Basic Working Package. The early detection of psychosocial problems belongs to the uniform part of that Working Package and should thus be offered to all children in a standardized way – as much as possible .

Differences between PCH professionals

In Chapter 2 we studied differences between individual PCH professionals in the number of children they identify as having problems. In order to do so data from three studies, using similar data collections methods and measurements were combined. We excluded children currently under treatment because of psychosocial problems, children from ethnic minorities and children with incomplete data. The remaining study sample consisted of over 3,000 children. These children had been examined by 117 PCH professionals in a routine health examination. Their parents had answered the Child Behavior Checklist (CBCL), a widely used and validated questionnaire on emotional and behavioral problems. On average PCH professionals identified 21% of all children as having problems. However, this percentage was shown to differ significantly between individual PCH professionals. Using multilevel analyses we determined whether these differences could be explained by differences in the prevalence of emotional an behavioral problems or in the presence of risk factors between the groups of children examined by each professional. We found that the differences between professionals could not be explained by these factors. In other words, whether a child with problems

was identified as such, did not only depend on the problems present but also on the individual PCH professionals by whom it was examined. This effect was greatest for children with a CBCL Total Problems Score (TPS) slightly above the clinical Cut-off point.

KIVPA

Chapter 3 assessed the psychometric qualities and added value of a short questionnaire on psychosocial problems of adolescents, the Korte Indicatieve Vragenlijst voor Psychosociale problematiek bij Adolescenten (KIVPA).

Based on the answers on the KIVPA a simple total sum score is calculated and three sub-scale scores. The sample consisted of over 1,200 pupils in second grade of secondary education, invited for a routine health examination by PCH. They answered the KIVPA and the Youth Self Report (YSR). Their parents answered the CBCL. Factor analyses were done to evaluate the scale structure. The results showed that the KIVPA, essentially, is a one-dimensional instrument. The hypothetical sub-scales were not replicated in the factor analyses. With the cut-off point recommended by the original authors, sensitivity for a clinical CBCL TPS was 0.57. That means that 57% of children with a clinical CBCL TPS, a strong indication of the presence of problems, was detected. Using the YSR and referral because of psychosocial problems as criteria, sensitivity was 0.82, respectively 0.55. The specificity for these criteria were 0.84, 0.85 and 0.83. The KIVPA was more sensitive for Internalizing than for Externalizing problems. The KIVPA was also found to offer added value: it allowed for a more accurate distinction between children with and without problems than was possible using only risk factors known to PCH.

We concluded that the KIVPA can be suitable for PCH, but that adaptations are desirable, in order to make the instrument more sensitive, especially for Externalizing problems[a].

LSPPK

In Chapter 4 we evaluated the psychometric properties of a short questionnaire on psychosocial problems for children aged 4 or 5, the Landelijk Signaleringsinstrument Psychosociale Problematiek Kleuters (LSPPK). The LSPPK is answered by parents. Their answers are summarized in a simple sum score, the Parent Index (PI). Following the routine health examination the PCH Professional indicates which problems are present in his or her opinion. Their answers are summarized in the Child Health Professional Index (CHPI). The sample consisted of over 1,200 parents of children invited for a routine health examination by PCH. Parents had answered the CBCL and the LSPPK. Factor analyses resulted in factors that closely corresponded to the theoretical scale structure. Using the cut-off point recommended by the developers of the questionnaires, the sensitivity for the criterion 'being treated because of psychosocial problems' was 0.50. In other words, the LSPPK detected 50% of all children being treated for such problems. The sensitivity for a clinical score on the CBCL TPS, a strong indication of the presence of problems, was 0.69. Using a lower

---

[a] In order to prevent misunderstanding: following this study the KIVPA was adapted. This new version has not yet been validated.

cut-off point would improve the sensitivity, but would also result in a steep increase of the percentage of children being 'detected' without reason (specificity). The CHPI did not result in a better detection. We also determined whether the questionnaire could offer added value to PCH: we assessed whether using the LSPPK allowed for a better distinction between children with and without problems, than a distinction based only on of risk factors, known to PCH. This proved to be the case.

We concluded that the LSPPK can improve the identification of children with psychosocial problems, but that the instrument needs to be improved[b].

PSC

In Chapter 5 the psychometric qualities and added value of the Pediatric Symptom Checklist (PSC) were evaluated. The PSC is a questionnaire for parents and was developed in the USA. Studies in the USA showed it to be a valid and suitable instrument for the detection of psychosocial problems. The concept of added value in this study was extended: we not only determined whether it could improve the identification using known risk factors, but also whether it improved the distinction based on the clinical judgment of PCH professionals after the routine health examination, not supported by validated questionnaires or other methods.

The sample consisted of 674 parents of children aged seven to twelve, invited for a routine health examination by PCH. They answered the PSC and the CBCL. The PSC allows for the calculation of a single sum score. PSC scores for Dutch children were lower than those for children in the USA. Therefore we used not only the cut-off point recommended by the authors, but also an adapted cut-off point. The internal consistency of the PSC was high (Cronbach's $\alpha = 0.89$). However, a confirmatory factor analysis showed that the single sum score was an inadequate description of the items and their interrelationships. The sensitivity for a clinical CBCL TPS was 0.72. Specificity was 0.93. Sensitivity for CBCL Internalizing and Externalizing problems were quite comparable and somewhat lower than for the CBCL TPS. Using 'being treated because of psychosocial problems' as criterion the sensitivity was 0.42, with a specificity of 0.90. The PSC was shown to offer a significant added value for PCH.

Nearly all parents were able to answer the PSC, but 20% made critical remarks on the questionnaire, especially concerning the discrepancies between questions and the answering categories offered.

It was concluded that the PSC is suitable for the identification of children with problems. To prevent missing too many children with problems, a lower cut-off point than the one recommended in the USA should be used.

Randomized comparison of PSC, SDQ and PSYBOBA

In Chapter 6 we compared the psychometric qualities and added value of three questionnaires: the PSC, the Strengths and Difficulties Questionnaire (SDQ) and the PSYBOBA The PSYBOBA is a questionnaire for parents of children aged 7 to 12 and was developed by the LSPPJ, specifically for Dutch PCH. The SDQ was developed in the United Kingdom and was proven to be suitable for the detection of children with

---

[b] Following this study, the LSPPK was also adapted. The adapted version has not yet been validated.

problems there. We used the SDQ Parent Form. The PSC is a questionnaire for parents and was developed in the USA. Studies in the USA showed it to be a valid and suitable instrument for the detection of psychosocial problems. The sample consisted of more than 2,000 parents of children invited for a routine health examination by PCH. They answered the CBCL and either the PSC or the PSYBOBA or the SDQ. Which parents got which of these three questionnaires was randomized, in order to enable a randomized comparison.

The scale structure of the questionnaires was evaluated with Cronbach's α and confirmatory factor analyses. The internal consistency of each of the questionnaires was high and – corrected for the number of items – identical. The confirmatory factor analyses showed that for each of the three questionnaires the scales were an inadequate description of the items and their interrelationships. The cut-off point recommended by the original authors resulted in large differences in sensitivity, while the Areas under Curve were quite similar. We therefore defined adapted cut-off points, those scores that resulted in a specificity of at least 0.90. Using these cut-offs the sensitivity for a clinical CBCL TPS, a strong indication of the presence of problems, varied between 0.78 (PSC) and 0.86 (PSYBOBA and SDQ). Differences were not statistically significant. Sensitivity for a borderline CBCL TPS was somewhat lower and, again, we found no significant differences between the questionnaires. Each of the three questionnaires was found to offer significant and substantial added value and thus allowed for a better identification by PCH of children with problems. The PSC performed somewhat less here, but differences were not statistically significant.

Most parents were able to answer the questionnaires. Parents rated the PSC somewhat more often as difficult. Twenty percent of the parents made critical remarks on the PSC, compared to ten percent for SDQ and the PSYBOBA. PCH professionals found using the SDQ more difficult, mainly because of the complexity of the calculation of sub-scales.

It was concluded that each of the three questionnaires could improve the identification by PCH of children with problems. Psychometric qualities do not offer definite arguments to prefer one of the three above the other ones.

Four SDQ based classification methods

In Chapter 7 a comparison is made between four different SDQ based methods to distinguish between children with and without problems. These classification methods are:

1. a normal vs. elevated SDQ TPS

2. parents reporting serious problems, yes or no

3. a normal vs. elevated score in one of the SDQ's problem sub-scales, in combination with an elevated score on the impairment scale

4. a combination of the methods 1, 2 and 3

The sample consisted of more than 700 parents of children aged seven to twelve, invited for a routine health examination by PCH. They answered the SDQ and the CBCL before the routine health examination. We determined the sensitivity and specificity of these four classification methods using 4 criteria: a clinical score on the CBCL TPS, on

CBCL Internalizing Problems, CBCL Internalizing problems and currently being treated because of psychosocial problems.

We found some differences between the four methods, but most of them were not statistically significant. The second method (parent-reported serious problems) was less sensitive for a clinical CBCL TPS, a strong indication of the presence of problems, than the combination method and this method is somewhat more sensitive for a clinical score on Internalizing Problems than method 3 (subscales). The first method did not perform significantly better or worse than the other methods, on each of the criteria used. Each method was shown to allow for a better discrimination between children with and without problems than identification based on clinical judgment and we found no significant differences in this respect.

We concluded that for a first identification of children with problems PCH could use the relatively simple first method: namely determine whether the child has an elevated score on the SDQ TPS.

Computer Adaptive Test

In Chapter 8 we investigated whether the items of the CBCL, SDQ, PSC and PSYBOBA could be used to develop an IRT-based CAT, allowing for an efficient yet accurate identification of children with problems. Using IRT it is possible to determine on which point of a supposed scale items are informative. Suppose, on a test for arithmetic ability, a child has answered the question "713 : 23 =??" correctly. In that case it will not be informative to ask it the question "6 : 3=??" too. Using this information one can select those items that are useful in an individual test administration: children with no or only less serious problems do not have to answer questions that are informative – i.e. distinctive – for children with serious problems and vice versa

The sample consisted of the same respondents as those in the study reported in Chapter 6. They were randomly divided into two subsamples. The analyses done in the first subsample proved that almost all items could be used in an IRT-based CAT. We developed such a CAT and used that for a simulation study in the second subsample: answers given by parents on the paper and pencil questionnaires were used as if they were given in response to CAT questions. De results showed that discrimination based on the CAT scores between children with and without problems was very accurate. Sensitivity for a clinical CBCL TPS, a strong indication of the presence of problems, was 0.89, with a specificity of 0.91. The CAT procedure was also shown to be effective. On average less than 12 items were needed to determine with 95% accuracy whether a child scored above or below the cut-off point. We concluded that an IRT-based CAT is a promising option for the identification of children with problems. The results of the simulation study, though, have to be replicated in real life, before this method can be implemented in PCH practice.

Discussion and implications

In Chapter 9 the three main research questions of this thesis were answered. The results were discussed and recommendations were made, both for PCH practice and for research.

The answer on the first research question, concerning differences between individual PCH professionals), is that there are indeed significant differences between these professionals and that these differences cannot be explained by differences in the prevalence of problems or risk factors between the groups examined by individual PCH professionals.

Concerning the second research question, we found that the short questionnaires we evaluated, can indeed improve the identification by PCH. The KIPVA and the LSPPK do so less well than the other questionnaires that we evaluated, the PSC, SDQ and PSYBOBA.

The third question concerned the possibility to develop an IRT-based CAT with which an efficient and accurate distinction between children with and without problems could be made. We showed that most items of the CBCL, SDQ, PSC and PSYBOBA were suitable for such a CAT. The simulation study offered strong evidence that such a CAT allows for an efficient and accurate identification of children with problems.

In the discussion we argued that the systematic differences between individual PCH professionals that we found indicate strongly that identification by PCH of children with problems based only on a clinical judgment does not meet the requirements of standardization that must be met for services belonging to the uniform part of PCH's Basic Working Package. Methods to improve this identification are therefore needed.

The methodological approach of the studies in this thesis allow for a valid generalization of our findings towards the large majority of children under PCH care. However, further research is needed among children from ethnic minorities.

We discussed the value of the criterion measures which were used in this thesis, the CBCL, YSR and being under treatment because of psychosocial problems. We argue that one cannot expect high associations between short questionnaires like the ones we studied and treatment status. The main argument for this is that many children with problems are not being treated. Based on the recommendations of an expert meeting in 1999, we argue that the CBCL and the YSR are relevant and well validated criterion measures for short questionnaires on psychosocial problems. However, they may not be regarded as real golden standards. This is one of the reasons that results of these short questionnaires must be interpreted carefully. Our findings, therefore, do not justify the use of the questionnaires as stand-alone screeners.

The CAT we developed appears a very promising method for the identification of children with problems. However, thus far the evidence for its validity and efficiency is based on a simulation study only. Before this CAT can be implemented, a study is needed assessing its validity and efficiency in real life, as well as its feasibility for parents and in the context of PCH practice.

Sometimes it is argued that the detection of psychosocial problems is only justified when there are sufficient and accessible services available for all problems identified. This proposition must be labeled as shortsighted, as well as in defiance with PCH's Basic Working Package. It is short sighted, because a high-quality identification will also result in the referral of less children without serious problems. That way, more resources will be available for those who need it most. It is in defiance with the Basic Working Package, because one of the PCH's tasks is, also, to collect information on the

population's health and to provide local and higher authorities with reliable policy information.

Younger children, up to 4 years of age, are seen regularly by PCH. Older children are seen less often. That means that problems developing between the consecutive routine health examinations will not be detected in an early stage. In the context of the so called school based ZATs (Zorg en Advies Teams, Care and Advice Teams) methods for continuous monitoring and detection can be developed and applied. PCH is expected to participate in those teams but these activities are not part of the uniform part of PCH's Basic Working Package. This anomaly should be redressed.

Finally we discuss the way short questionnaires are actually being used in PCH practice. The studies presented in this thesis showed that these questionnaires can improve the identification by PCH. However, the extent to which they will actually do so, not only depends on the quality of the questionnaires, but also on the way they are used in practice. Conversations with PCH professionals and impressions from PCH practice lead us to believe that there are large differences between individual professionals and between individual PCH services. Therefore a careful implementation and continuous quality management is needed.

Implications

Based on the results and the discussion the following recommendations concerning PCH practice are made:

1. For the identification of children with problems PCH should not rely on the clinical judgment of individual professionals alone, but should use methods and instruments which have been shown to be valid.
2. It should use these instruments as they were validated and for the purpose for which they were evaluated; changes should only be allowed after new validation studies.
3. The introduction of instruments in PCH practice should be carefully planned and should be embedded in a continuous implementation and quality management program, ensuring that all staff members know how to use them and continue to use them as intended.
4. A lack of adequate youth healthcare services should be considered as a major argument in favor of high-quality methods to identify children with problems, not used as an argument against it.
5. The identification of children with problems in the ZAT-context must be seen as an essential addition to their identification in the routine health examination. The participation of PCH in these ZATs must be redefined and included in the Uniform Part of the Basic Working Package.

For future research the following recommendations are made:

1. Methods and instruments need to be developed for all groups for which such instruments are not yet available and/or validated, especially children under 5 and adolescents over 14.
2. The quality of the identification of children with problems among ethnic minorities needs to be assessed.
3. Instruments and methods need to be developed and validated which allow for a systematic assessment by PCH of the nature and seriousness of problems and the

need for professional help, when short questionnaires indicate the likelihood of problems.

4. Methods need to be developed to collect data from more than one informant; especially since the Internet and computer adaptive testing may be promising methods in this respect.

5. The question to what extent and under which conditions short questionnaires do actually improve the identification of and care for children with problems in daily practice needs to be addressed.

6. The validity and feasibility of a Computer Adaptive Test in day-to-day PCH practice needs to be investigated.

7. Evaluating the relative merits of a number of comparable questionnaires is to be preferred over the evaluation of a single questionnaire.

11 Samenvatting

Dit proefschrift behandelt de signalering door de Jeugdgezondheidszorg (JGZ) van kinderen met psychosociale problemen en gaat daarbij met name in de op de vraag of korte vragenlijsten die signalering kunnen verbeteren. De vragen die in de verschillende hoofdstukken beantwoord werden, zijn:

1. Er zijn verschillen tussen individuele JGZ-medewerkers in het aantal door hen gesignaleerde kinderen met problemen. Zijn die verschillen significant en kunnen zij verklaard worden door verschillen in de prevalentie van problemen of risicofactoren bij de door hen onderzochte kinderen?

2. Wat zijn de psychometrische eigenschappen van een aantal korte vragenlijsten die de JGZ zou kunnen gebruiken bij die signalering en kunnen zij die signalering verbeteren?

3. Is het mogelijk om op basis van Item Response Theory (IRT) een Computer Adaptive Test (CAT) te ontwikkelen om daarmee op een efficiënte wijze een nauwkeurige indicatie van de aanwezigheid van problemen te krijgen?

In navolging van een expert meeting in 1999 wordt de term psychosociale problematiek in dit proefschrift gebruikt om emotionele en gedragsproblemen en problemen op school aan te duiden.

Inleiding

In het eerste hoofdstuk werd beschreven hoe de JGZ in de jaren zeventig van de vorige eeuw uitdrukkelijker aandacht ging besteden aan psychosociale problematiek. Aanvankelijk waren de mate waarin en de wijze waarop dat gebeurde vooral afhankelijk van lokale interesses. Dat leidde tot een grote heterogeniteit in de aanpak. De inzet van de Landelijke Werkgroep Signalering van Psychosociale Problematiek bij Jongeren (LSPPJ) en de invoering van het Basistakenpakket JGZ veranderden dat. Signalering van kinderen met psychosociale problemen behoort tot het uniforme deel van het Basistakenpakket (Monitoring en Signalering) en dient – voor zover mogelijk – aan alle kinderen op een uniforme wijze aangeboden te worden.

Verschillen tussen JGZ-medewerkers

In hoofdstuk 2 werd een onderzoek beschreven naar verschillen tussen individuele JGZ-medewerkers in het aantal kinderen met problemen dat zij signaleren. Om dat te kunnen doen werden de gegevensbestanden van een drietal onderzoeken met vergelijkbare dataverzamelingsmethodieken en vergelijkbare gegevens samengevoegd. Na uitsluiting van kinderen die momenteel onder behandeling zijn voor psychosociale problemen, van kinderen met een allochtone herkomst en van kinderen met onvolledige gegevens, beschikten we over gegevens van ruim 3.000 kinderen. Deze kinderen waren door 117 verschillende JGZ-medewerkers onderzocht in het kader van het standaard Periodiek Gezondheidsonderzoek (PGO). Na afloop daarvan gaven de JGZ-medewerkers hun oordeel over de aanwezigheid van psychosociale problemen. De ouders zelf vulden voorafgaand aan het PGO de Child Behavior Checklist (CBCL) in, een veel gebruikt en ruim gevalideerd meetinstrument voor emotionele en gedragsproblemen. Gemiddeld signaleerden JGZ-medewerkers bij bijna 21% enige psychosociale problematiek. Dat percentage varieerde sterk tussen individuele JGZ-medewerkers en die verschillen bleken groter dan op basis van toeval verwacht mocht worden. Met behulp van multilevel analyses onderzochten we of die verschillen verklaard konden worden door

verschillen in prevalentie van problemen of risicofactoren tussen de groepen kinderen die door individuele medewerkers onderzocht waren. Dat was niet het geval. Met andere woorden: of een kind met problemen werd gesignaleerd, hing niet alleen af van de aanwezige problematiek, maar ook van de vraag door welke JGZ-medewerker het desbetreffende kind werd onderzocht. Dat speelde met name bij kinderen met een score op de CBCL die iets boven het klinische afkappunt lag.

KIVPA

In hoofdstuk 3 werden de psychometrische kwaliteiten en de toegevoegde waarde van de Korte Indicatieve Vragenlijst voor Psychosociale problematiek bij Adolescenten (KIVPA) onderzocht. De KIVPA is een korte vragenlijst, bedoeld voor adolescenten. Op basis van de antwoorden op de KIVPA wordt een totale somscore berekend, plus scores op drie subschalen. De onderzoeksgroep bestond uit ruim 1200 leerlingen in de tweede klas van het voortgezet onderwijs. Zij vulden de KIVPA en de Youth Self Report (YSR) in. Hun ouders beantwoordden de CBCL.

De veronderstelde schaalstructuur werd onderzocht met behulp van factoranalyses. De resultaten lieten zien dat de KIVPA in essentie een eendimensionaal instrument is; de veronderstelde subschalen werden door de factoranalyses niet gerepliceerd. Met het aanbevolen afkappunt was de sensitiviteit voor een klinische CBCL TPS, een sterke aanwijzing voor de aanwezigheid van problemen, 0,57. Dat wil zeggen dat 57% van de kinderen met een klinische CBCL TPS werd gedetecteerd. Met de YSR en met verwijzing op grond van psychosociale problemen was de sensitiviteit 82% respectievelijk 55%. De specificiteit bij deze criteria was respectievelijk 0,84, 0,85 en 0,83. De LSPPK was sensitiever voor Internaliserende problemen dan voor Externaliserende problemen. Ook de KIVPA bleek een toegevoegde waarde te kunnen bieden aan de JGZ, in die zin dat het instrument een beter onderscheid mogelijk maakt dan met bij de JGZ bekende risicofactoren alleen zou kunnen.

Geconcludeerd werd dat de KIVPA een geschikt instrument lijkt voor de JGZ, maar dat aanpassing wenselijk lijkt om het instrument sensitiever te maken, met name voor Externaliserende Problemen[c].

LSPPK

In hoofdstuk 4 werden de psychometrische eigenschappen van het Landelijk Signaleringsinstrument Psychosociale Problematiek Kleuters (LSPPK) onderzocht. Ook werd nagegaan of de LSPPK voor de JGZ toegevoegde waarde biedt: we onderzochten of het instrument een beter onderscheid tussen kinderen met en zonder problemen mogelijk maakte dan dat op basis van bekende risicofactoren zou kunnen. De LSPPK is een korte vragenlijst bedoeld voor ouders van kleuters. Op basis van de door ouders aangegeven problemen wordt een somscore berekend, in dit proefschrift PI genoemd (Parent Index). Na afloop van het PGO geeft de JGZ-medewerker aan welke problemen er volgens hem aanwezig zijn (CHPI, Child Health Professional Index). De onderzoeksgroep bestond uit ruim 1200 kinderen waarvan de ouders voorafgaand aan het PGO de LSPPK en de CBCL beantwoord hadden. De resultaten van factoranalyses

---

[c] Om misverstanden te voorkomen zij hier benadrukt dat de KIVPA na het hier gerapporteerde onderzoek is aangepast. Die aanpassing is nog niet gevalideerd.

kwamen goed overeen met de veronderstelde schaalstructuur. Bij het door de ontwikkelaars van de LSPPK aanbevolen afkappunt voor de PI was de sensitiviteit voor het criterium 'onder behandeling zijn voor psychosociale problemen' 0,50. Met andere woorden, de LSPPK detecteerde 50% van de kinderen die voor psychosociale problemen onder behandeling zijn. De sensitiviteit voor een score boven het klinische afkappunt van de CBCL Totale Probleem Score (TPS), een sterke aanwijzing voor de aanwezigheid van problemen, was 0.69. Verlaging van het afkappunt resulteerde weliswaar in meer kinderen met problemen die gedetecteerd worden, maar het percentage kinderen dat ten onrechte wordt gedetecteerd (specificiteit) werd dan snel groter. De CHPI resulteerde niet in een betere signalering. De LSPPK bleek een beter onderscheid tussen kinderen met en zonder problemen mogelijk te maken dan op basis van bij de JGZ bekende risicofactoren mogelijk was.

Geconcludeerd werd dat de LSPPK de signalering kan verbeteren, maar ook dat het instrument verbetering behoeft.[d]

PSC

In hoofdstuk 5 werden de psychometrische eigenschappen en de meerwaarde van de Pediatric Symptom Checklist (PSC) onderzocht. De PSC is een vragenlijst voor ouders, afkomstig uit de Verenigde Staten en is daar valide en bruikbaar gebleken voor de signalering van psychosociale problemen bij kinderen. Het begrip meerwaarde werd in dit hoofdstuk uitgebreid ten opzichte van de wijze waarop het in hoofdstuk 3 en 4 werd gebruikt: er werd niet alleen nagegaan of het instrument de signalering op basis van risicofactoren zou kunnen verbeteren, maar ook of het de signalering door de JGZ op basis van het periodiek onderzoek – niet ondersteund door een vragenlijst – zou kunnen verbeteren. De onderzoeksgroep bestond uit 674 ouders van kinderen van 7 tot 12 jaar, uitgenodigd voor het PGO. Zij vulden de PSC in en de CBCL. Op basis van de antwoorden op de PSC werd een enkele totale probleemscore berekend. De scores op de PSC voor de Nederlandse kinderen bleken significant lager dan die voor kinderen in de VS. Daarom werd bij de analyses niet alleen het oorspronkelijke afkappunt, maar ook een aangepast afkappunt gehanteerd. De interne consistentie van de PSC TPS bleek hoog ($\alpha = 0,89$). Niettemin bleek uit een toetsende factoranalyse dat de eendimensionale PSC TPS de items in hun onderlinge samenhang niet adequaat weergaf. De sensitiviteit voor een klinische CBCL TPS, een sterke aanwijzing voor de aanwezigheid van problemen, was bij het aangepaste afkappunt 0,72; de specificiteit was 0,93. De sensitiviteit voor Internaliserende problemen en voor Externaliserende problemen waren vergelijkbaar, maar lagen wat lager dan voor een klinische CCBCL TPS. Met het criterium 'nu onder behandeling zijn' waren sensitiviteit en specificiteit respectievelijk 0,42 en 0,90. De PSC bleek de signalering op basis van risicofactoren en het normale PGO significant te kunnen verbeteren en heeft dus een duidelijke toegevoegde waarde voor de JGZ.

---

[d] Ook de LSPPK is op basis van de onderzoeksresultaten inmiddels aangepast. De validiteit van de aangepaste LSPPK is tot op heden niet onderzocht.

152

De PSC kon door vrijwel alle ouders ingevuld worden, maar 20% maakte kritische opmerkingen over de vragenlijst, vooral over de discrepantie tussen vragen en de geboden antwoorden.

Geconcludeerd werd dat de PSC bruikbaar is voor de signalering van kinderen met problemen. Om te voorkomen dat daarbij te veel kinderen met problemen worden gemist, moet een lager afkappunt dan in de VS gehanteerd worden.

Gerandomiseerde vergelijking van PSC, SDQ en PSYBOBA

In hoofdstuk 6 werden de psychometrische eigenschappen en de meerwaarde voor de JGZ van drie vragenlijsten met elkaar vergeleken: de PSC, de vragenlijst voor PSYchosociale problematiek in de BOvenbouw van het BAsisonderwijs (PSYBOBA) en de Strengths & Difficulties Questionnaire (SDQ). De PSYBOBA werd ontwikkeld door de Landelijke Werkgroep Signalering van Psychosociale Problematiek bij Jongeren (LSPPJ) voor gebruik door de JGZ en is bedoeld voor ouders van kinderen van 7 tot 12. De SDQ is afkomstig uit het Verenigd Koninkrijk en is daar valide en bruikbaar gebleken voor de signalering van kinderen met problemen. De PSC is een vragenlijst voor ouders, afkomstig uit de Verenigde Staten en is daar valide en bruikbaar gebleken voor de signalering van psychosociale problemen bij kinderen.

De onderzoeksgroep bestond uit meer dan 2000 ouders van kinderen. Zij vulden de CBCL in en óf de PSC óf de PSYBOBA óf de SDQ. Welke ouder welke van deze vragenlijsten kreeg, was gerandomiseerd om op die manier een gerandomiseerde vergelijking van de eigenschappen van de lijsten mogelijk te maken. De schaalstructuur van de lijsten werd onderzocht met behulp van Cronbach's α en met toetsende factoranalyse. De interne consistentie van de drie lijsten was goed en – gecorrigeerd voor het aantal items – identiek. De toetsende factoranalyse toonde aan dat de schalen bij elke lijst een onvoldoende beschrijving gaven van de items in hun onderlinge samenhang. De door de auteurs aanbevolen afkappunten resulteerden in grote verschillen in sensitiviteit en specificiteit, terwijl de Area Under Curves zeer vergelijkbaar waren. Daarom werden aangepaste afkappunten gedefinieerd, namelijk dat punt dat resulteerde in een specificiteit van 0,90. Met die afkappunten bleek de sensitiviteit voor een klinische CBCL TPS, een sterke aanwijzing voor de aanwezigheid van problemen, te variëren tussen 0,78 (PSC) en 0,86 (PSYBOBA en SDQ). De verschillen in sensitiviteit waren niet significant. De sensitiviteit voor een borderline CBCL TPS was wat lager, maar ook hier waren de verschillen niet statistisch significant. Elk van de lijsten bleek de signalering door de JGZ op basis van risicofactoren en het PGO substantieel en significant te kunnen verbeteren. Bij de PSC was dat wat minder dan bij de andere lijsten, maar het verschil was opnieuw niet significant.

De meeste ouders konden elk van de lijsten goed invullen. Ouders vonden de PSC wat vaker moeilijk. Twintig procent van de ouders maakte kritische opmerkingen over de PSC, tegenover tien procent bij de PSYBOBA en de SDQ. JGZ-medewerkers vonden de SDQ moeilijker om te gebruiken, met name door de complexiteit van de berekening van subschalen.

Geconcludeerd werd dat elk van de lijsten de signalering door de JGZ zou kunnen verbeteren. Psychometrische eigenschappen vormden geen doorslaggevend argument voor de keuze voor een van de onderzochte instrumenten.

Vier classificatiemethoden van de SDQ

In hoofdstuk 7 werden vier verschillende methoden vergeleken om op basis van de SDQ een onderscheid te kunnen maken tussen kinderen met en zonder problemen, namelijk:

1. een al dan niet verhoogde SDQ TPS

2. ouders rapporteren ernstige problemen, ja versus nee

3. een al dan niet verhoogde score op een van de drie probleem-subschalen van de SDQ, in combinatie met een verhoogde score op de impairment-schaal

4. een combinatie van de methoden 1, 2 en 3.

De onderzoeksgroep bestond hier uit 711 ouders van kinderen die werden uitgenodigd voor een PGO. Voorafgaand aan het PGO vulden zij de CBCL en de SDQ in.

We bepaalden de sensitiveit en specificiteit van de vier classificatiemethoden voor vier criteria: een klinische score op de CBCL TPS, CBCL Internaliserende problemen, CBCL Externaliserende problemen, en onder behandeling zijn voor psychosociale problemen.

Er waren een paar verschillen tussen de vier methoden, maar de meeste waren statistisch niet significant. De tweede methode (door ouders gerapporteerde problemen) was minder sensitief voor een klinische CBCL dan de combinatiemethode, en de combinatiemethode is wat sensitiever voor een klinische score voor Internaliserende problemen dan methode 3 (verhoogde subschaal). De eerste methode (een verhoogde TPS) was niet significant beter of slechter dan de andere methodes, ongeacht het criterium. Elke methode kon de signalering door de JGZ op basis van het klinisch oordeel na het PGO verbeteren en zij verschilden in dit opzicht niet significant van elkaar.

Geconcludeerd werd dat voor een eerste identificatie van kinderen met problemen de JGZ de relatief eenvoudige, eerste methode kan gebruiken: namelijk bepalen of de score op de SDQ TPS boven of onder het afkappunt valt.

Computer Adaptive Test

In hoofdstuk 8 werd nagegaan of de items van de CBCL, SDQ, PSC en PSYBOBA gebruikt konden worden om met behulp van Item Response Theory (IRT) een Computer Adaptive Test (CAT) te maken waarmee op een efficiënte manier een nauwkeurige indicatie verkregen kan worden van aanwezige problematiek.

IRT maakt het mogelijk om te bepalen op welk punt van een (veronderstelde) schaal items informatief zijn. Stel dat een kind in een rekentoets het goede antwoord geeft bij de som "713 : 23 =?". Het is dan niet informatief meer om het ook de som "6 : 3=?" voor te leggen. Op die manier kan een selectie gemaakt worden van items die in een individuele afname zinvol zijn: kinderen met weinig problemen hoeven geen items voorgelegd te krijgen die informatief – want onderscheidend – zijn voor kinderen met veel problemen, en andersom.

De onderzoeksgroep bestond uit dezelfde kinderen als in hoofdstuk 6. De respondenten werden at random verdeeld over twee subsamples. De uitgevoerde IRT-analyses op het eerste sample wezen uit dat vrijwel alle items voor een IRT-based CAT gebruikt zouden kunnen worden. Met het tweede sample deden we een simulatieonderzoek. We ontwierpen een CAT en gebruikten de antwoorden op de papieren vragenlijsten als waren het antwoorden op die CAT. De resultaten van die CAT bleken een goed onderscheid te kunnen maken tussen kinderen zonder en met problemen. De sensitiviteit voor een klinische CBCL TPS was 0,89, met een specificiteit van 0,91. Dat gebeurde ook op een efficiënte manier: gemiddeld waren minder dan 12 items nodig om met 95% nauwkeurigheid te bepalen of kinderen onder of boven het afkappunt scoorden.

Geconcludeerd werd dat een op IRT gebaseerde CAT een veelbelovende optie is voor de identificatie van kinderen met problemen. De bevindingen van het simulatieonderzoek moeten wel nog in de JGZ-praktijk bevestigd worden.

Discussie en implicaties

In hoofdstuk 9 ten slotte werden antwoorden gegeven op de geformuleerde onderzoeksvragen, werden de bevindingen besproken en werden aanbevelingen geformuleerd voor verder onderzoek en voor de JGZ-praktijk.

Het antwoord op de eerste onderzoeksvraag (verschillen tussen individuele medewerkers) is dat er inderdaad significante verschillen zijn tussen individuele JGZ-medewerkers en dat die niet verklaard kunnen worden door verschillen in problemen en risicokenmerken in de door hen onderzochte groepen.

De tweede onderzoeksvraag luidde of korte vragenlijsten de signalering door de JGZ kunnen verbeteren. We hebben aangetoond dat de onderzochte vragenlijsten de signalering door de JGZ inderdaad kunnen verbeteren. De KIVPA en de LSPPK doen dat echter bij kleuters en adolescenten minder goed dan de SDQ, PSYBOBA of PSC bij kinderen van 7 tot 12.

De derde onderzoeksvraag betrof de mogelijkheid om met behulp van IRT een CAT te ontwerpen die met weinig vragen een nauwkeurige indicatie van aanwezige problematiek zou moeten opleveren. We hebben aangetoond dat de meeste items van de CBCL, SDQ, PSC en PSYBOBA gebruikt kunnen worden voor zo'n CAT. Het simulatieonderzoek gaf ook sterke aanwijzingen dat met een dergelijke CAT op een zeer efficiënte wijze een zeer valide indicatie van aanwezige problematiek verkregen kan worden.

In de discussie van de bevindingen werd betoogd dat de aangetoonde stelselmatige verschillen tussen individuele JGZ-medewerkers een sterke aanwijzing vormen dat de signalering door de JGZ op basis van een klinisch oordeel na het PGO en zonder gebruik van gevalideerde methoden niet voldoet aan de eis van standaardisatie die gesteld wordt aan het uniforme deel van het basistakenpakket JGZ. Methodieken om die signalering te verbeteren zijn daarom noodzakelijk.

De wijze waarop de onderzoeken zijn uitgevoerd impliceren dat de bevindingen geldig geacht mogen worden voor het grootste deel van de kinderen die bij de JGZ onder zorg zijn, maar dat voor generalisatie naar kinderen uit etnische minderheden nader onderzoek noodzakelijk is.

155

De waarde van de in dit proefschrift gehanteerde criteriummaten (met name CBCL, YSR en onder behandeling zijn) werd aan de orde gesteld. Beargumenteerd werd dat niet verwacht kan worden dat er een zeer hoge correlatie zal bestaan tussen signaleringslijsten en het criterium 'onder behandeling zijn'. Dit omdat uit onderzoek voldoende is gebleken dat lang niet alle kinderen die ernstige problemen hebben feitelijk bij de hulpverlening terecht komen. In aansluiting op de aanbevelingen van de eerder genoemde expertmeeting werd betoogd dat de CBCL en de YSR inhoudelijk relevante en goed gevalideerde criteriummaten zijn voor signaleringslijsten voor psychosociale problematiek. Wel kunnen zij niet gezien worden als een echte gouden standaard. Mede daarom, moeten bevindingen met de onderzochte signaleringslijsten ook altijd met zorg geïnterpreteerd worden. Het verrichte onderzoek is zeker niet te beschouwen als een validatie van het gebruik van de onderzochte signaleringslijsten als preselectie-instrumenten.

De ontwikkelde CAT lijkt een zeer veelbelovende methode te zijn voor signalering van psychosociale problematiek, maar de aanwijzingen voor efficiëntie en validiteit zijn gebaseerd op simulatieonderzoek, niet op een echte afname van de CAT in de praktijk. Voordat de CAT breed geïmplementeerd kan worden, moeten de bruikbaarheid, validiteit en efficiëntie in een onderzoek in de JGZ-praktijk nader onderzocht worden.

In de discussie werd ook ingegaan op de opvatting dat signalering van psychosociale problemen alleen dan zinvol en legitiem zou zijn als er voldoende, goed toegankelijke voorzieningen zijn voor alle gesignaleerde problemen. Die stelling werd als kortzichtig bestempeld en als strijdig met het takenpakket van de JGZ in Nederland. Kortzichtig, omdat een goede signalering er ook aan bijdraagt dat alleen die kinderen die echt hulp nodig hebben naar voorzieningen verwezen worden. Strijdig met de taken van de JGZ, omdat het de uitdrukkelijke taak van de JGZ is informatie over de gezondheid van de bevolking te verzamelen om op die manier lokale en hogere overheden van beleidsinformatie te voorzien.

De JGZ ziet jonge kinderen regelmatig. Bij wat oudere kinderen is dat veel minder vaak het geval. Daardoor kunnen problemen die tussen de verschillende contactmomenten ontstaan moeilijker vroegtijdig gesignaleerd worden. In het kader van de ZorgAdviesTeams in het onderwijs kan gewerkt wordt aan methodieken voor continue monitoring en signalering. De JGZ wordt geacht daarin te participeren, maar die activiteiten zijn momenteel geen onderdeel van het uniforme deel van het Basistakenpakket. Die tegenstrijdigheid moet opgelost worden.

Tot slot werd het feitelijk gebruik van vragenlijsten aan de orde gesteld. De onderzoeken in dit proefschrift hebben laten zien dat vragenlijsten de signalering door de JGZ kunnen verbeteren. De mate waarin dat feitelijk het geval is, is niet alleen afhankelijk van de vragenlijsten zelf, maar ook van de wijze waarop zij gebruikt worden. Uit gesprekken met JGZ-medewerkers bestaat de indruk dat er op dit punt forse verschillen bestaan. Dat vraagt om een zorgvuldige implementatie en een voortdurend kwaliteitsbeleid op dit punt.

Implicaties

De bevindingen en de discussie leidden tot de volgende aanbevelingen voor de praktijk van de JGZ in Nederland:

1. De JGZ kan bij het signaleren van kinderen met problemen niet vertrouwen op het klinisch oordeel van individuele medewerkers, maar moet methodieken en instrumenten gebruiken waarvan de validiteit aangetoond is.

2. De JGZ moet die instrumenten gebruiken zoals zij gevalideerd zijn en voor het doel waarvoor zij gevalideerd zijn; veranderingen mogen pas doorgevoerd worden na nieuw validatieonderzoek.

3. De invoering van instrumenten in de JGZ moet zorgvuldig gepland worden en moet ingebed zijn in een voortdurend programma gericht op kwaliteitsbeheer, zodat alle medewerkers weten hoe de instrumenten te gebruiken en de instrumenten ook blijven gebruiken zoals bedoeld.

4. Een gebrek aan adequate voorzieningen voor hulpverlening moet gezien worden als een belangrijk argument voor de wenselijkheid van een goede signalering van kinderen met problemen, niet als een argument daartegen.

5. Signalering van kinderen met problemen in de context van ZorgAdviesTeams (ZAT) moet beschouwd worden als een essentiële aanvulling op de signalering in de context van het PGO; participatie van de JGZ in de ZAT moet dan ook gedefinieerd worden als een onderdeel van het Uniforme deel van het Basistakenpakket JGZ.

Voor toekomstig onderzoek worden de volgende aanbevelingen geformuleerd:

1. Methoden en instrumenten voor de signalering van psychosociale problemen moeten ontwikkeld worden voor alle groepen waarvoor zij nu nog niet beschikbaar zijn, met name voor kinderen jonger dan 5 en bij adolescenten ouder dan 14.

2. De kwaliteit van de signalering bij kinderen uit etnische minderheden moet vastgesteld worden.

3. Er moeten instrumenten ontwikkeld worden waarmee de JGZ aard en ernst van aanwezige problemen op een systematische wijze kan vaststellen, waneer korte signaleringslijsten wijzen op de waarschijnlijke aanwezigheid van problemen.

4. Er moeten methoden ontwikkeld worden om informatie te verzamelen bij meerdere informanten, ook omdat internet en computer adaptive testing veelbelovende opties zijn in dit opzicht.

5. De vraag in welke mate en onder welke condities korte signaleringslijsten de vroegsignalering in de praktijk daadwerkelijk verbeteren, moet beantwoord worden.

6. De validiteit en de bruikbaarheid van een Computer Adaptive Test in de alledaagse JGZ praktijk moet onderzocht worden.

7. Een vergelijkende evaluatie van de kwaliteiten van vragenlijsten verdient de voorkeur boven de evaluatie van een enkele vragenlijst.

Appendix 1   List of abbreviations

ADHD	Attention-Deficit/Hyperactivity Disorder
AUC	Area under Curve
BJZ	Bureau Jeugdzorg
CAT	Computerized Adaptive Testing
CB	Consultatiebureau
CBCL	Child Behavior Checklist
CFA	Confirmatory Factor Analysis
CHC	(Preventive) Child Healthcare; elsewhere: PCH
CHP	(Preventive) Child Health Professional
CHPI	Child Health Professional Index
CI	Confidence Interval
DAWBA	Development and Well-being Assessment
Df	Degree of freedom
GGD	Gewestelijke/Gemeentelijke GezondheidsDienst
GGD-NL	GGD Nederland
IRT	Item Response Theory
ITSEA	Infant and Toddler Emotional Assessment Scale
JGZ	Jeugdgezondheidszorg
KIVPA	Korte Indicatieve Vragenlijst voor Psychosociale problematiek bij Adolescenten
LSPPJ	Landelijke werkgroep Signaleringsinstrumenten Psychosociale Problematiek Jeugd
LSPPK	Landelijke Signaleringshulp Psychosociale Problematiek Kleuters
NIPG	Nederlands Instituut voor Praeventieve Gezondheidszorg
NIZW	Nederlands Instituut voor Zorg en Welzijn
NVAZ	Nederlandse Vereniging voor AdolescentenZorg
NVJG	Nederlandse Vereniging voor Jeugdgezondheidszorg
OECD	Organisation for Economic Co-Operation and Development
OR	Odds ratio
P	Probability
PC	Principal Component
PCFI	Parsimony Corrected Fit Index
PCH	Preventive Child Healthcare; elsewhere: CHC
PGO	Periodiek GezondheidsOnderzoek
PI	Parent Index
PSC	Pediatric Symptom Checklist
PSYBOBA	Vragenlijst voor PSYchosociale problematiek in de BOvenbouw van het BAsisonderwijs
RIV	Random Intercept Variance
ROC	Receiver Operating Characteristics
Sd	Standard deviation
SDQ	Strengths & Difficulties Questionnaire
SEM	Structural Equation Modeling
SPSS	Statistical Package for the Social Sciences
TNO	Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek

TPS	Total Problem Scale
TDS	Total Difficulties Score
US	United States
VPC	Variance Partition Component
Wajong	Wet Arbeidsongeschiktheidsvoorziening jonggehandicapten
YSR	Youth Self Report
ZonMw	Nederlandse Organisatie voor gezondheidsonderzoek en zorginnovatie

Appendix 2 Graduate School for Health Research SHARE

This thesis is published within the research program **Public Health and Public Health Services Research** of the Graduate School for Health Research SHARE. More information regarding the institute and its research can be obtained from our internet site: www.rug.nl/share.

Previous dissertations from the program **Public Health and Public Health Services Research**

Akker-Scheek I van den (2007) *Recovery after short-stay total hip and knee arthroplasty; evaluation of a support program and outcome determination*
Supervisors: prof dr JW Groothoff, prof dr SK Bulstra
Co-supervisors: dr M Stevens, dr W Zijlstra

Mei S van der (2007) *Social participation after kidney transplantation*
Supervisors: prof dr WJA van den Heuvel, prof dr JW Groothoff, prof dr PE de Jong
Co-supervisor: dr WJ van Son

Khan MM (2007) *Health policy analysis: the case of Pakistan*
Supervisors: prof dr WJA van den Heuvel, prof dr JW Groothoff
Co-supervisor: dr JP van Dijk

Rosenberger J (2006) *Perceived health status after kidney transplantation*
Supervisors: prof dr JW Groothoff, prof dr WJA van den Heuvel
Co-supervisors: dr JP van Dijk, dr R Roland

Šléškova M (2006) *Unemployment and the health of Slovak adolescents*
Supervisors: prof dr SA Reijneveld, prof dr JW Groothoff
Co-supervisors: dr JP van Dijk, dr A Madarasova-Geckova

Dumitrescu L (2006) Palliative care in Romania
Supervisor: prof dr WJA van den Heuvel

The B (2006) *Digital radiographic preoperative planning and postoperative monitoring of total hip replacements; techniques, validation and implementation*
Supervisors: prof dr RL Diercks, prof dr JR van Horn
Co-supervisor: dr ir N Verdonschot

Jutte PC (2006) *Spinal tuberculosis, a Dutch perspective; special reference to surgery*
Supervisor: prof dr JR van Horn
Co-supervisors: dr JH van Loenhout-Rooyackers, dr AG Veldhuizen

Leertouwer H (2006) *Het heil van de gezonden zij onze hoogste wet; de geschiedenis van de medische afdeling bij de arbeidsinspectie*
Supervisors: prof dr JW Groothoff, prof dr MJ van Lieburg, prof dr D Post

Jansen DEMC (2006) *Integrated care for intellectual disability and multilpe sclerosis*
Supervisors: prof dr D Post, prof dr JW Groothoff
Co-supervisor: dr B Krol

Ham I van (2006) *De arbeidssatisfactie van de Nederlandse huisarts*
Supervisors: prof dr J de Haan, prof dr JW Groothoff
Co-supervisor: dr KH Groenier

Jansen GJ (2005) *The attitude of nurses towards inpatient aggression in psychiatric care: the development of an instrument*
Supervisors: prof dr SA Reijneveld, prof dr ThWN Dassen
Co-supervisor: dr LJ Middel

Post M (2005) *Return to work in the first year of sickness absence; an evaluation of the Gatekeeper Improvement Act*
Supervisors: prof dr JW Groothoff, prof dr D Post
Co-supervisor: dr B Krol

Landsman-Dijkstra JJA (2005) *Building an effective short healthpromotion intervention; theory driven development, implementation and evaluation of a body awareness program for chronic a-specific psychosomatic symptoms*
Supervisor: prof dr JW Groothoff
Co-supervisor: dr R van Wijck

Bakker RH (2005) De samenwerking tussen huisarts en bedrijfsarts
Supervisor: prof dr JW Groothoff
Co-supervisors: dr B Krol, dr JWJ van der Gulden

Nagyová I (2005) *Self-rated health and quality of life in Slovak rheumatoid arthritis patients*
Supervisor: prof dr WJA van den Heuvel
Co-supervisor: dr JP van Dijk

Gerritsma-Bleeker CLE (2005) *Long-term follow-up of the SKI knee prosthesis.*
Supervisors: prof dr JR van Horn, prof dr RL Diercks
Co-supervisor: dr NJA Tulp

Vries M de (2005) *Evaluatie Zuidoost-Drenthe HARTstikke goed!; mogelijkheden van community-based preventie van hart-en vaatziekten in Nederland*
Supervisors: prof dr D Post, prof dr JW Groothoff
Co-supervisor: dr JP van Dijk

Jungbauer FHW (2004) *Wet work in relation to occupational dermatitis*
Supervisors: prof dr PJ Coenraads, prof dr JW Groothoff
Post J (2004) Grootschalige huisartsenzorg buiten kantooruren
Supervisor: prof dr J de Haan

Reneman MF (2004) *Functional capacity evaluation in patients with chronic low back pain; reliability and validity*
Supervisors: prof dr JW Groothoff, prof dr JHB Geertzen
Co-supervisor: dr PU Dijkstra

Bâra-Ionilã C-A (2003) *The Romanian health care system in transition from the users' perspective*
Supervisors: prof dr WJA van den Heuvel, prof dr JAM Maarse
Co-supervisor: dr JP van Dijk

Lege W de (2002) *Medische consumptie in de huisartspraktijk op Urk*
Supervisors: prof dr D Post, prof dr JW Groothoff

Hoekstra EJ (2002) *Arbeidsbemiddeling met behulp van Supported Employment als interventie bij de reïntegratie van chronisch zieken; de rol van de arbeidsbemiddelaar, chronisch zieke en werkgever*
Supervisors: prof dr JW Groothoff, prof dr K Sanders, prof dr WJA van den Heuvel, prof dr D Post

Enk JG van (2002) *Determinants of use of healthcare services in childhood*
Supervisors: prof dr D Post, prof dr AJP Veerman, prof dr WJA van den Heuvel
Gecková A (2002) Inequality in health among Slovak adolescents
Supervisors: prof dr D Post, prof dr JW Groothoff.
Co-supervisor: dr JP van Dijk

Dijk JP van (2001) *Gemeentelijk gezondheidsbeleid; omvang en doelgerichtheid*
Supervisors: prof dr D Post, prof dr M Herweijer, prof dr JW Groothoff

Middel LJ (2001) *Assessment of change in clinical evaluation*
Supervisor: prof dr WJA van den Heuvel
Co-supervisor: dr MJL de Jongste

Bijsterveld HJ (2001) *Het ouderenperspectief op thuiszorg; wensen en behoeften van ouderen ten aanzien van de thuis(zorg)situatie in Friesland*
Supervisors: prof dr D Post, prof dr B Meyboom-de Jong
Co-supervisor: dr J Greidanus

Dijkstra GJ (2001) *De indicatiestelling voor verzorgingshuizen en verpleeghuizen*
Supervisors: prof dr D Post, prof dr JW Groothoff

Dalen IV van (2001) *Second opinions in orhopaedic surgery: extent, motives, and consequences*
Supervisors: prof dr JR van Horn, prof dr PP Groenewegen, prof dr JW Groothoff

Beltman H (2001) *Buigen of barsten? Hoofdstukken uit de geschiedenis van de zorg aan mensen met een verstandelijke handicap in Nederland 1945-2000*
Supervisors: prof dr D Post, prof dr AThG van Gennep

Pal TM (2001) *Humidifiers disease in synthetic fiber plants: an occupational health study*
Supervisors: prof dr JGR de Monchy, prof dr D Post, prof dr JW Groothoff

Goossen WTF (2000) *Towards strategic use of nursing information in the Netherlands*
Supervisors: prof dr WJA van den Heuvel, prof dr ThWN Dassen, prof dr ir A Hasman

Hospers JJ (1999) *Allergy and airway hyperresponsiveness: risk factors for mortality*
Supervisors: prof dr D Post, prof dr DS Postma, prof dr ST Weiss

Wijk P van der (1999) *Economics: Charon of Medicine?*
Supervisors: prof dr WJA van den Heuvel, prof dr L Koopmans, prof dr FFH Rutten
Co-supervisor: dr J Bouma

Dijkstra A (1998) *Care dependency: an assessment instrument for use in long-term care facilities*
Supervisors: prof dr WJA van den Heuvel, prof dr ThWN Dassen

Tuinstra J (1998) *Health in adolescence: an empirical study of social inequality in health, health risk behaviour and decision making styles*
Supervisors: prof dr D Post, prof dr WJA van den Heuvel
Co-supervisor: dr JW Groothoff

Mink van der Molen AB (1997) *Carpale letsels: onderzoek naar de verzuimaspecten ten gevolgen van carpale letsels in Nederland 1990-1993*
Supervisors: prof dr PH Robinson, prof WH Eisma
Co-supervisors: dr JW Groothoff, dr GJP Visser

Mulder HC (1996) *Het medisch kunnen: technieken, keuze en zeggenschap in de moderne geneeskunde*
Supervisor: prof dr WJA van den Heuvel

Dekker GF (1995) *Rugklachten-management-programma bij de Nederlandse Aardolie Maatschappij B.V.: ontwerp, uitvoering en evaluatie*
Supervisors: prof dr D Post, prof WH Eisma
Co-supervisor: dr JW Groothoff

Puttiger PHJ (1994) *De medische keuring bij gebruik van persluchtmaskers*
Supervisors: prof dr D Post, prof dr WJA Goedhard
Co-supervisor: dr JW Groothoff

Engelsman C & Geertsma A (1994) *De kwaliteit van verwijzingen*
Supervisors: prof dr WJA van den Heuvel, prof dr FM Haaijer-Ruskamp,
prof dr B Meyboom-de Jong

Lucht F van der (1992) *Sociale ongelijkheid en gezondheid bij kinderen*
Supervisor: prof dr WJA van den Heuvel
Co-supervisor: dr JW Groothoff