Self-report in Youth Health Monitoring

Petra van de Looij-Jansen

Self-report in Youth Health Monitoring

Evidence from the Rotterdam Youth Monitor

Petra van de Looij-Jansen

Colofon

ISBN 978-94-6050-001-5

Lay out and cover design: Marco van Bemmel, Vorm en Ontwerp
Printed by Ipskamp Drukkers, Enschede

The research projects were performed within the framework of CEPHIR, the academic workplace of Erasmus MC, department of Public Health, and the Municipal Public Health Service Rotterdam-Rijnmond.

Financial support by the Municipal Public Health Service Rotterdam-Rijnmond, the department of Public Health, Erasmus MC, the Erasmus University Rotterdam, GGD Nederland and the Jurriaanse Stichting for printing this thesis is gratefully acknowledged by the author.

Self-report in Youth Health Monitoring

Evidence from the Rotterdam Youth Monitor

Zelfrapportage in jeugdgezondheidsonderzoek

Kennis uit de Jeugdmonitor Rotterdam

Proefschrift

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

Prof.dr. H.G. Schmidt

en volgens besluit van het College van Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 1 april 2010 om 11.30 uur

door

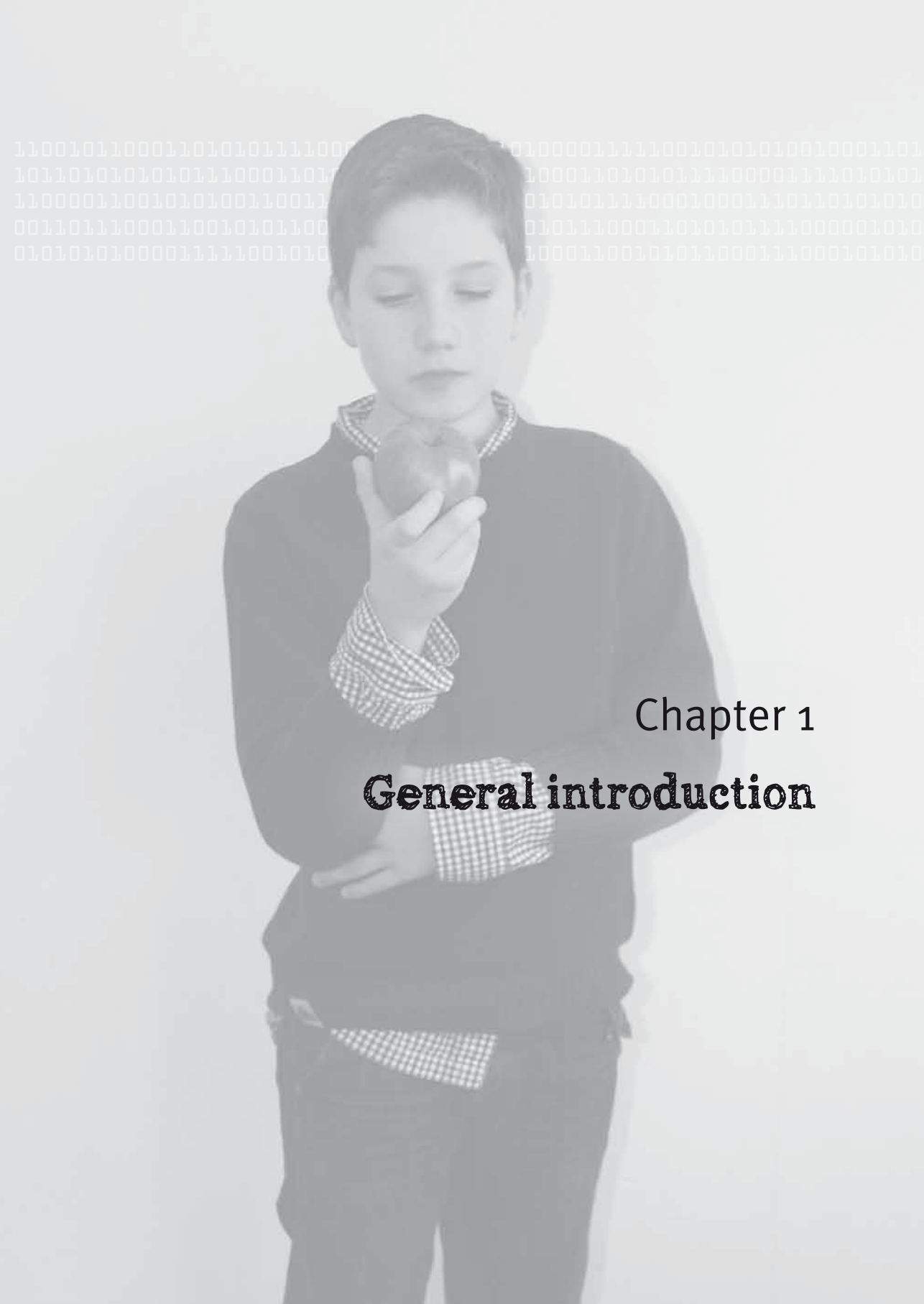
Petra Monique van de Looij-Jansen

geboren te Arkel

Promotiecommissie

Promotor:	Prof.dr. M.C.H. Donker
	Prof.dr. F.C. Verhulst
Overige leden:	Prof.dr. P. Muris
	Prof.dr. R.A. HiraSing
	Dr. H. Raat
Copromotor:	Dr. E.J. de Wilde

Contents



Chapter 1

General introduction

Under Dutch law, preventive youth healthcare organisations have a duty to ensure the early identification of children with health or developmental problems. Similarly, municipalities have a duty to monitor young people's health at least every four years. For problem identification and monitoring, both individual and collective, these organisations often use self-report questionnaires. The overall aim of this thesis is to study various methodological and validity issues related to the use of self-report questionnaires among young people in a preventive youth healthcare setting. Seven specific research questions are derived from the Rotterdam Youth Monitor (RYM), a longitudinal youth health surveillance system integrated into preventive youth healthcare in the greater Rotterdam area.

Perspectives on validity of self-report questionnaires

Consider the following question:

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain or the item seems daft! Please give your answers on the basis of how things have been for you over the last six months.

	Not True	Somewhat True	Certainly True
I am often unhappy, down-hearted or tearful	☐	☐	☐

It is not hard to imagine that responses to this question (item 13 from the Strengths and Difficulties Questionnaire, (Goodman, 2001)) will greatly depend on the respondent's mood and circumstances at the time of answering. The influence of mood is unlikely to be problematic, as the question is intended to assess an aspect of mood (along with some of the other items treated in the overall questionnaire). But the influence of circumstances most certainly *is* problematic. Despite the clear instruction stated above the question, it is very possible, for example, that a student who has just discovered that her boyfriend has cheated on her will not assess her mood over the last six *months* but over the last six *hours*. And even if her mental review of the last six months leads her to conclude that she has actually been unhappy during that period, she will not fill in the correct answer if the questionnaire is being completed in a classroom situation in which the person at the next desk is openly discussing her answers with other members of the class.

The notion that research participants may misrepresent their behaviour to researchers has its historical roots in the theoretical work of Goffman (1959). Goffman's idea of the 'presentation of self in everyday life' suggests that individuals consciously portray themselves in a manner that conforms to socially normative expectations.

This has clear implications for any research into human behaviour, even non-obtrusive research. Whether or not they know they are being observed, people tend to represent themselves in this way – a tendency that may be even more pronounced when a researcher asks them to report facts, feelings and behaviours. The tendency is probably greatest when the questions concern sensitive or threatening topics, such as abuse, criminal behaviour, alcohol and drug use, and psychological well-being.

If respondents alter their responses to conform to norms of socially acceptable behaviour, measurement bias is inevitable: they may underreport behaviours they perceive to be socially unacceptable – behaviours which, if detected by others, would result in disapproval or punishment.

To explain the source of validity problems that may emerge with some self-reported data, two major theoretical perspectives have been advanced: 1) the cognitive perspective and 2) the situational perspective (Brener et al., 2003). These can be summarised as follows:

1) The cognitive perspective

By focusing on the mental processes underlying self-reported data, this perspective attributes validity problems to inaccuracies arising from comprehension, recall, and other cognitive operations.

The basic models proposed for the cognitive process (Cannell et al., 1981; Jobe and Herrmann, 1996; Tourangeau et al., 2000) postulate that the question-answering process is influenced by four basic cognitive processes: a) comprehension of the question; b) recall of requested information from memory; c) decision-making; and d) response generation. Potentially, errors can arise at each of these stages, thus contributing to validity problems. An additional problem is that validity may vary from question to question, as the specific cognitive operations used in responding to a question may differ according to factors such as the length of the reference period and the type of response required. For example, a question may ask about the frequency of a certain behaviour, or merely whether this behaviour occurred.

In general, the more complex the recall task, the less reliable the reporting. This suggests that the quality of responses can be improved by using strategies designed to enhance recall, such as relatively short reference periods, and by formulating the questions in simple language. The way in which various cognitive processes are executed may also be influenced by the context of the question, including the other questions in the questionnaire and the environment in which the questionnaire is administered. This in turn can lead to systematic biases in responding.

2) The situational perspective

This perspective focuses on validity problems that are not the product of cognitive processing but of various characteristics of the external environment. Here, two factors are presumed to be particularly influential: the presence of other people while a respondent is answering questions, and respondents' perceptions of the degree of privacy or confidentiality that their responses will receive (Tourangeau and Yan, 2007).

Various constructs are used to explain situational biases. One of them is social desirability: the desire to provide others with a favourable impression of oneself. Response biases may also result from a fear of reprisal, which can be caused by a perceived lack of confidentiality, anonymity, or privacy in the situational context. Such concerns are particularly likely to lead to underreporting on behaviours that are illegal, stigmatized, or laden with moral implications. Increasing empirical evidence does indeed suggest that respondents' willingness to report sensitive information or illegal activities is influenced by the mode of interview used in self-report data collection (Aquilino, 1994; Gfroerer and Hughes, 1991; Sudman and Bradburn, 1982).

The critical factor appears to be response anonymity during the interview: the degree to which an interview mode allows respondents to record their answers without revealing potentially embarrassing information to an interviewer or to others nearby (Aquilino, 1994; Bradburn, 1983). This explains the advantage of self-administered questionnaires over telephone and face-to-face modes: they provide greater response anonymity. As Gfroerer showed, if adolescent respondents do not perceive the interview environment to be private and confidential, they report significantly lower rates of alcohol and drug use (Gfroerer, 1985).

The use of self-report questionnaires in youth health surveys

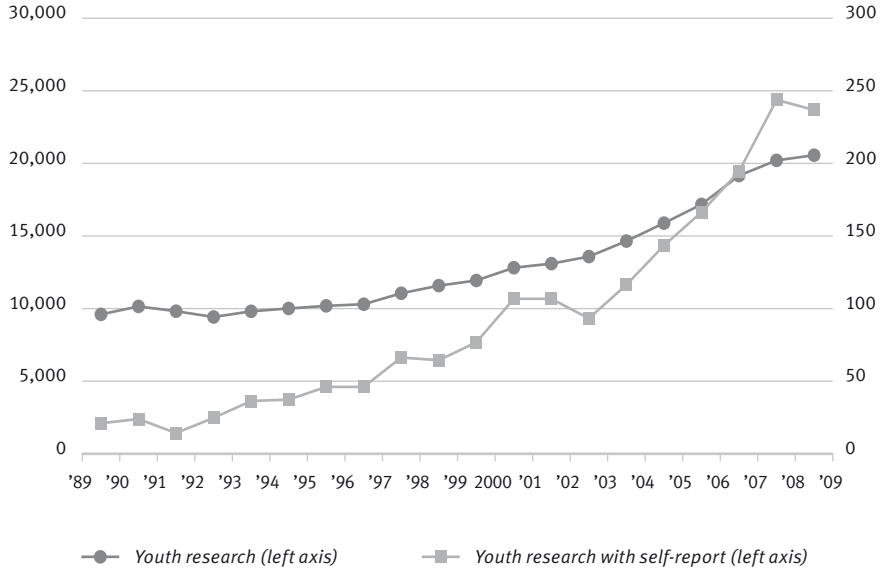
Self-report questionnaires now play an accepted role not only in youth research, but also in youth policy, as demonstrated by UNICEF's recent international comparative studies on child well-being (Bradshaw et al., 2007). In the recent development of international Child Indicators (as described by the International Society of Child Indicators, *www.childindicators.org*), self-reported indicators of well-being are now preferred above other means of measurement.

Such questionnaires are used for a variety of reasons. First, many behaviours (such as unprotected sexual intercourse or the use of alcohol or illicit drugs) are far more difficult to investigate through observation. Similarly, indicators of psychological well-being (such as self-esteem, personal efficacy, depression, or health-related quality of life) are almost impossible to observe: interview or self-report are nearly the only ways in which they can be measured.

It is therefore no surprise that the number of youth studies (published in peer-reviewed journal articles) using self-report questionnaires for some of their data collection has substantially increased over the last three decades. Figure 1 illustrates this increase: while 0.2% of all journal articles published in 1989 had used self-report questionnaires, the figure was 1.6% for those published in 1990.[1]

[1] *Search in PubMed. Search strings for youth research: (adolescence or adolescents or youth or child or childhood) in title, limits Humans, All Infant: birth-23 months, All Child: 0-18 years, Newborn: birth-1 month, Infant: 1-23 months, Preschool Child: 2-5 years, Child: 6-12 years, Adolescent: 13-18 years. Search strings for youth research with questionnaires: ((self-report or self-completed) and (questionnaire or survey)) and (adolescence or adolescents or youth or child or childhood) in title, limits Humans, All Infant: birth-23 months, All Child: 0-18 years, Newborn: birth-1 month, Infant: 1-23 months, Preschool Child: 2-5 years, Child: 6-12 years, Adolescent: 13-18 years.*

Figure 1.1 The number of youth studies published in peer-reviewed journals with self-report measures.



This increase is not due solely to methodical advantages: it also reflects a growing interest in youth research and youth monitoring. Because overweight and depression in youth have become major public health issues, an increasing number of studies have used self-report to investigate their prevalence and determinants. Similarly, the need for brief, well-validated screening instruments in the study of mental health problems has led to the constant development and testing of these instruments' psychometric properties (for example the Strengths and Difficulties Questionnaire) (Vogels, 2008).

Another reason for the increasing use of self-report is that automated methods for data-entry (such as data-scanning) and data-collection (such as internet-based questionnaires) now make it much easier to conduct youth surveys.

A third reason is unique to the Netherlands: the fact that local governments are legally obliged to monitor residents' health (Public Health Act, 2008). To this end, all municipal public health services conduct health surveys at least every four years. Naturally, these include surveys of young people's health.

Context of the Rotterdam Youth Monitor (RYM)

The use of self-report questionnaires has also increased in preventive youth healthcare. In the Netherlands, the youth health care system plays an important role in the early identification of physical, psychosocial and child-raising problems. This is because all children aged 0-19 are invited for routine health examinations (approximately fifteen times between 0-4, and three times at primary and secondary school (Boudewijnse et al., 2005)).

To ensure that each child in the Netherlands is offered the same minimum level of services, basic services of preventive youth health care are defined in the Basic Working Package for Health Care in 0-19-year-olds (*Basistakenpakket Jeugdgezondheidszorg 0-19 jaar*)(Ministry of Health, Welfare and Sports, 2002), which has been laid down by law (Ministry of Health, Welfare and Sports, 2008).

As part of the Basic Working Package, certain standardised services are systematically provided for all children. The services in the Basic Working Package comprise: 1) the identification and monitoring of developmental problems; 2) the estimation of care needs; 3) screening and immunization; 4) health promotion, advice, instruction and support (for purposes of improving health-related behaviour); 5) action to avert health risks; 6) support to the healthcare system. The first of these services (the monitoring and identification of problems) is provided at two levels: the individual level, in order to monitor a child's development and early identify problems; and the collective level, in order to monitor young people's overall health by developing health profiles intended to inform local policymakers and stimulate evidence-based public health. Under the 2008 Public Health Act, the latter has to be carried out every four years.

Although meta-analysis by Durlak and Wells (1997) showed that early identification improves the prognosis of children, several studies have indicated the need to improve the early identification of psychosocial problems (emotional, behavioural and educational problems) in preventive youth health care. For instance, the study by Brugman et al. (2001) showed that 43% of the children whose total score on the Child Behaviour Check List (CBCL, parent report) indicated the need for treatment (Verhulst et al., 1996) were not identified by a physician or nurse. Similar results were found by Reijneveld et al. (2004).

One strategy for improving the quality of early identification in preventive youth health care involves using validated questionnaires to help identify psychosocial problems. In 2008, national recommendations were made in the Netherlands on the use of questionnaires for this purpose, though validated instruments are not yet available for all age groups (Postma, 2008). A particularly promising questionnaire is the Strengths and Difficulties Questionnaire (SDQ), whose parent version has been validated in the Netherlands for the 7-12 age group (Goodman, 2001; Vogels et al., 2005). Validation studies for the other age groups are still taking place.

In the context of preventive youth healthcare, the Municipal Public Health Service for the Rotterdam area runs the Rotterdam Youth Monitor, a longitudinal youth-health surveillance system that monitors the general health, well-being, behaviour and related factors of the

0-19 age group in the greater Rotterdam area (*www.jeugdmonitorrotterdam.nl*). Introduced in 1996, the RYM has become a major instrument both for preventive youth healthcare and for local youth policy. It serves to inform youth policy at school, neighbourhood, and municipality levels. Because its data are integrated into the routine health examinations provided by the preventive youth health care system to five age groups (2, 5-6, 10-11, 12-13, and 14-15), they are also used to detect potential individual health risks or problems. This supports any necessary preventive measures that may be necessary, including referrals for treatment.

Most RYM data are collected through self-report questionnaires. For children below the age of 10, however, reports by parents and teachers are used; for children aged 10-11, self-reports are combined with parents' and teacher's reports. From 12 years onwards, only adolescents' self-reports of adolescents are used. Because an overall summary of respondents' answers (marked on a form called 'points of special interest'; 'Aandachtspuntenformulier' in Dutch) is an input into the routine health examinations by school nurses, all data are collected according to a confidential procedure. Data are collected throughout the year, except in July and August. The questionnaires focus mainly on emotional and behavioural problems, but also include several correlates of emotional and behavioural problems. Every two years, reports are made for all participating schools; every four years, reports are made for municipalities and neighbourhoods.

Aim and outline of this thesis

As stated at the beginning of this introduction, the use of self-report questionnaires raises several validity issues. This thesis addresses a number of these issues.

More specifically, it studies a number of methodological and validity issues related to the Rotterdam Youth Monitor. These issues are divided into two parts.

The first part examines the extent to which the reports provided by respondents to self-report questionnaires depend on the circumstances under which the questionnaires were administered. These concern the different methods of data collection (i.e. anonymous versus confidential; and web-based versus paper-and-pencil-based). This part of the study also covers the situational perspective – aspects such as social desirability, privacy, and confidentiality – as well as seasonal variations in self-reported health and health-related behaviour.

The second set of issues concerns the validity of measures used in the RYM: the validity of self-reported length and weight, and the factorial validity of the SDQ in various groups of young people, the predictive validity of discrepancies between parent-reported and child-reported emotional problems, and the predictive validity of emotional problems (including suicide) for the various actions taken by school nurses.

Yet more specifically, the thesis answers the following research questions (see also overview studies in Table 1.1):

Table 1.1 Overview of the different studies presented in this thesis

Chapter	Focus	Study population	N
	Method of data collection		
2	Anonymous versus confidential	grade 5 secondary school	704
3	Web versus paper-and-pencil	grade 3 secondary school	532
4	Seasonal variation	grade 1 and 3 secondary school	33,171
	Validity of self-report measures		
5	Validity of self-report of height and weight	grade 1 secondary school	499
6	Factorial validity and factorial invariance self-report SDQ	grade 1 and 3 secondary school	11,881
7	Predictive validity of parent-child disagreement of self-reported internalizing problems	grade 6 primary school and grade 1 secondary school	1,170
8	Predictive validity of self-reported suicidal thoughts and behaviour	grade 1 secondary school	3,692

Method of data collection

1. Do anonymous self-report questionnaires on health and health-related behaviour produce different results than confidential ones?
2. Do web-based self-report questionnaires on health and health-related behaviour produce different results than paper-and-pencil ones?
3. Do self-reports on health and health-related behaviour differ according to the season in which a questionnaire is administered?

Validity of self-report measures

4. Are self-reported length and weight valid tools for estimating the prevalence of overweight and obesity?
5. What is the factorial validity of the self-report Strengths and Difficulties Questionnaire? Is the factor structure of this questionnaire invariant across gender, age, ethnic background, and level of education?
6. What is the predictive validity of disagreement in internalizing problems between preadolescent children and their parents?
7. What is the predictive validity of self-reported suicidal behaviour in adolescents?

The final chapter contains a general discussion of the findings presented in this thesis, and addresses directions for future research and implications for public health practice.

References

Aquilino WS (1994), Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly* 54: 362-395.

Boudewijnse HB, Lokven v, E., Oskam Ek eds. (2005), *Nederlands Leerboek Jeugdgezondheidszorg deel A en deel B*. Assen: Koninklijke Van Gorcum BV.

Bradburn N (1983), Response effects. In: *Handbook of survey research (pp. 289-328)*, Rossi P, Wright J, Anderson A, eds. San Diego, CA: Academic.

Bradshaw J, Hoelscher P, Richardson D (2007), *'Comparing Child Well-being in OECD Countries: Concepts and Methods'* Innocenti Working Paper (No. 2006-03). Florence: UNICEF Research Centre.

Brener ND, Billy JOG, Grady WR (2003), Assessment of factors affecting the validity of self-reported health-risk behaviour among adolescents: Evidence from the scientific literature. *J Adolesc Health* 33: 436-457.

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

Cannell CF, Miller PV, Oksenberg L (1981), Research on interviewing techniques. In: *Sociological Measurement*, (ed) LS, ed. San Francisco: Jossey-Bass, pp 389-437.

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

Gfroerer JC (1985), Influence of privacy on self-reported drug use by youths. In: *Self-report methods of estimating drug use: meeting current challenges to validity* Rockville, MD: U.S. Department of Health and Human Services, pp 22-30.

Gfroerer JC, Hughes AL (1991), The feasibility of collecting drug abuse data by telephone. *Public Health Rep* 106: 384-393.

Goffman E (1959), *The presentation of self in everyday life* New York: Doubleday.

Goodman R (2001), Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry* 40: 1337-1345.

Jobe JB, Herrmann D (1996), Implications of models of survey cognition for memory theory. In: *Basic and applied memory research: Volume 2: Practical applications*, D. Herrmann MJ, C. McEvoy, C. Herzog, P. Hertel (Eds), ed. Hillsdale, NJ: Erlbaum, pp 193-205.

Ministry of Health, Welfare and Sports (2002), *Basistakenpakket jeugdgezondheidszorg 0-19 jaar.* The Hague: Ministry of Health, Welfare and Sports.

Ministry of Health, Welfare and Sports (2008), Public Health Act. Sdu Publishers, The Hague.

Postma S (2008), *JGZ-richtlijn. Vroegsignalering van psychosociale problemen*. Bilthoven: RIVM, Centrum Jeugdgezondheidszorg.

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

Sudman S, Bradburn NM (1982), *Asking Questions.* San Francisco: Jossey-Bass.

Tourangeau R, Rips L, Rasinski K (2000), *The psychology of survey response* Cambridge, UK: Cambridge University Press.

Tourangeau R, Yan T (2007), Sensitive questions in surveys. *Psychol Bull* 133: 859-83.

Verhulst FC, Van der Ende J, Koot HM (1996), *Handleiding voor de CBCL/4-18 [Manual for the CBCL/4-18]* Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

Vogels AGC (2008), *The identification by Dutch Preventive Child Health Care of children with psychosocial problems: do short questionnaires help?* Thesis.

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

Chapter 2

Comparison of anonymous versus confidential survey procedures:

effects on health indicators in Dutch adolescents

Petra M. van de Looij-Jansen, Judith E.J. Goldschmeding, Erik Jan de Wilde

Journal of Youth and Adolescence 2006; 35 (4): 659-665

Abstract

Self-report questionnaires are frequently used in youth research, but doubt remains whether total anonymity affects the results. This study compared the responses of 704 mainly 16-17-year-old adolescents to self-report measures of various health indicators in two groups: anonymous and confidential collection. For most health indicators no significant differences were found. Good perceived health, use of alcohol, and aggressive behaviour (for boys) were reported significantly less frequently in the confidential group (small effect size). It appeared that pupils with high scores on social desirability do not respond different at most health indicators in the two conditions. For most health characteristics, epidemiologically useful outcomes may be derived from confidential self-report measures, which are also valuable for detecting individual at-risk children. Similar studies in lower grades and other types of education should confirm the results.

Introduction

Self-report questionnaires have acquired an accepted role in youth research. They are easy to administer in a class-room setting, can address a wide range of topics that would otherwise require lengthy interviews, and usually require no complicated techniques of analysis. In general, the use of self-report instruments seems to eliminate variation attributable to interviewer characteristics (Tourangeau et al., 1997). Also, respondents are much more likely to disclose potentially embarrassing information when the questions are self-administered than when they are administered by an interviewer (Tourangeau et al., 2000).

With the increase of their use, the variety of methods of administration has increased. Some studies use questionnaires in a practically anonymous fashion, with no names on the questionnaire or questions within that could lead to identification. Other questionnaires are distributed with names and dates of birth on the cover, as they are part of a longitudinal design or serve as a preparation for later research. Mail surveys may be sent out to a specific person and returned anonymously. Whoever actually fills out the questionnaire and in what condition remains unclear. Recently, electronic questionnaire administration has added an extra dimension in the variety.

Although it could be expected that the method of assessment influences response, especially for questions on those kind of behaviours that seem to correlate negatively with social desirability, studies so far on this topic are not conclusive. In one of the earliest studies, Kulik et al. (1968) investigated whether anonymity is needed for a truthful self-report of antisocial behaviour. A checklist of antisocial behaviours was filled out both anonymously and non-anonymously by high schools and delinquent boys. Both groups disclosed more antisocial behaviour when protected by anonymity, but differences were comparatively small. Moreover, such differences were greater for minor violations (such as drinking alcohol to excess), than for serious ones (striking a teacher). They concluded that the importance of anonymity has been overemphasized in research on delinquency using self-report techniques. Similar studies were carried out on other subjects, mainly on substance use. King (1970) compared anonymous with precoded drug usage questionnaires in a mailed survey of undergraduates and found no statistically significant differences between the two groups with respect to number and percentage of returns, or reports of use of marijuana and LSD. Malvin and Moskowitz (1983) investigated differences between anonymous versus identifiable self-reports of drug attitudes, intentions and use in junior high school students. Students in the identifiable condition reported less current use of cocaine than students in the anonymous condition. No effect was found for use of alcohol, barbiturates (tranquilizers), amphetamines (stimulants), marijuana (for girls) and cigarettes (for boys). The report of drug use and attitudes on drugs were not different in a study with a similar design by O'Malley et al. (2000). Bjarnason and Adalbjarnardottir (2000) found only small differences in a study with a corresponding design conducted among tenth graders.

Self-evidently, however, the sensitivity of a particular question is of importance. Thus, although Ong and Weiss (2000) found no differences in positive responses to questions about marijuana use, they did find differences to the questions about masturbation, cheating during exams, and stealing. The difference in response may be a result of cognitive factors such as the expectation that results will not be treated confidentially, despite the promise of the administering staff. Also, results may be influenced by the intention of the subject to depict a certain image of him - or herself - to the person who in the end sees the results. Intuitively, this *social desirability* would play a bigger part in confidential than in anonymous surveys.

In the Netherlands many municipal health services regularly use questionnaires on a variety of topics. A recent change in method in the Rotterdam preventive youth healthcare system raised questions about the implications of this change for epidemiological results: the outcome of individual questionnaires became part of the contents of an interview with the school nurse. This implied a change from anonymous self-report to a confidential procedure. As mentioned before, no consistent pattern appears from earlier studies. Therefore, the main research question is: Are there differences in the report of (psychological) health indicators between anonymous and confidential self-report questionnaires? Since there are substantial differences between boys and girls in health indicators in adolescence, we will study the influence of gender in this respect. Furthermore, the effect of social desirability, as one of the sources of response bias, will be taken into account in the analyses.

Methods

Study design

This study was conducted in seven secondary schools in Rotterdam City, with all pupils in the eleventh grade of general secondary education and pre-university education being approached for participation. The response rate of the pupils was 83%. A within-schools design was chosen which means that half of the classes were randomly assigned (by computer) to receive questionnaires with students' names on them (confidential group), while the other half of the classes were given questionnaires with no identifying information (anonymous condition). The questionnaires were administered by specially trained school nurses of the Municipal Public Health Service.

Nineteen classes were entered into the confidential group, with pupils being assured that all information provided was strictly confidential. They were told that completed questionnaires would "never be shown to teachers, parents, police or anyone else". To emphasize confidentiality, pupils in the identifiable group sealed the questionnaire in a blank envelope. Twenty classes were entered into the anonymous group, pupils in this condition being told that their answers would be anonymous - that is, that no one would be able to identify them personally.

The questionnaire was filled out in the classroom in about one hour. Teachers were present during the administration for keeping order in the classroom. Data were collected in November/December 2002.

Subjects

704 pupils participated in this study. Table 2.1 shows how various demographic variables were distributed between anonymous and confidential groups. Despite the random assignment, significant differences were found between the anonymous and confidential condition for level of education (χ^2=163.0, p<0.001) and age (F=8.40, p<0.01). In the anonymous group 76% of the pupils followed higher general secondary education, compared to 28% in the confidential group. Some 46% of the pupils were 16 years or younger in the anonymous group, compared to 56% in the confidential group.

Table 2.1 Demographic variables by condition

	anonymous condition		confidential condition	
	n	%	n	%
Gender				
boys	165	47.0%	165	46.7%
girls	186	53.0%	188	53.3%
Level of education**				
general secondary education	264	75.6%	97	27.5%
pre-university education	85	24.4%	256	72.5%
Age*				
15 years	9	2.6%	2	0.6%
16 years	151	42.9%	192	54.9%
17 years	134	38.1%	124	35.4%
18 years	52	14.8%	31	8.9%
19 years and older	6	1.7%	1	0.3%

*$p < 0.01$ **$p < 0.001$

Measures

The questionnaire of the Rotterdam Youth Monitor was used for this study. This questionnaire, developed for pupils attending secondary schools, consists of indicators on well-being, problem behaviour, health risk behaviour and some related factors (stressful life events, school factors). The following indicators were used for this study:

Self-esteem was measured by the Dutch version of the Rosenberg Self-esteem Scale (Rosenberg, 1965; van der Linden, 1983). This scale consists of 10 statements (e.g. "I feel that I have a number of good qualities"). Each item was answered on a 4-point scale (strongly disagree, disagree, agree and strongly agree). Scores range from 0 to 10, with a high score indicating higher self-esteem (k=10, α=0.89). Suicidal ideation was assessed with one item: "In the last 12 months, have you thought about putting an end to your life?" (Garnefski and Diekstra, 1993). Respondents answered this question on a 5-point scale (1=never, to 5=very often). Suicide attempt was measured by the following question: "Did you ever seriously attempt putting an end to your life?" (three answering categories: never, once, more than once). Mental health was measured by the 12-item version of the General Health Questionnaire (GHQ) (Koeter and Ormel, 1991). The GHQ is a self-report questionnaire consisting of statements about aspects of well-being, such as worries, tension or sleep loss. With each statement, the current status of the respondent (over the past four weeks) is compared with his or her 'normal' status by one of four responses. Scores range from 0 to 12, with a higher score indicating lower mental health (k=12, α=0.82). Perceived health was assessed by means of one

question in which the respondents were asked to rate their health (CBS, 2001). Respondents reported their answers on a 5-point scale ranging from very good (1) to bad (5).

The questionnaire included 14 items about aggressive and delinquent behaviour, developed and tested by the scientific research Department of the Ministry of Justice (Junger-Tas et al., 1992). Factor analysis of these items revealed three factors: 'aggressive behaviour' consisting of items about beaten up somebody, molested somebody at the street or joined in a fight (k=4, α=0.84), 'vandalism and stealing' consisting of items about damaging something at the street, set fire to something and stealing (k=6, α=0.85) and 'severe violent delinquent behaviour' consisting of more severe acts like beaten up somebody so badly that the victim needed a doctors treatment, threatened somebody to get money, threatened somebody with a weapon, injured somebody with a weapon and burglary (k=4, α=0.92). All items were answered on a 5-point scale ranging from never (0) to very often (4). Problem behaviour was also operationalised by one question about carrying a weapon (1=never to 3=every day), one question about playing truant (1=never to 6=more than 20 hours) and one question about bullying (1=never to 5=daily).

The following health risk behaviours were included in this study: smoking, alcohol consumption, soft drug use and hard drug use. These questions were based on a questionnaire used by the Netherlands Institute of Mental Health and Addiction (De Zwart, 1999). Smoking habits were measured by the question: "Have you ever smoked cigarettes?", with five possible answers, ranging from 'I have never smoked' (1) to 'I smoke daily' (5). For alcohol consumption we used the question concerning frequency of consumption in the past four weeks. Possible answers ranging from '0 times' (1) to 'more than 10 times' (6). Soft drug use was measured by the question "How often have you used cannabis in the past four weeks?", with possible answers ranging from '0 times' (1) to 'more than 10 times' (6). Hard drug use was assessed by four questions: use of xtc, use of coke, use of speed and use of horse with possible answers ranging from '0 times' (1) to 'more than 10 times' (6). A sum score of these four questions was calculated (0-20).

The questionnaire included eleven questions about stressful life events, for example 'dead family member', 'psychological problems parent', 'divorce parents' and 'sexual abuse'. 'Problems at home' was assessed by calculating a sum score of five items about the home situation, for example 'are there many tensions at home?', 'do you often feel alone at home?', and 'did you think about running away from home?' (α=0.74). Scores range from 0 to 5, with a higher score indicating more problems at home. School factors include perception of academic functioning which is measured by a sum score of 6 questions (α=0.80), perception of relationship with classmates which is measured by a sum score of seven questions (α=0.80), and perception of relationship with teachers which is measured by a sum score of six questions (α=0.72). Scores range from 0 to 100, with a higher score indicating a more positive perception of academic functioning, a more positive perception of relationship with classmates or a more positive perception of relationship with teachers (Stoel, 1980). Victim of bullying at school was measured by one question asking about the frequency of being bullied in the past

year, with answering categories 'never' (1), 'incidental', 'once a week', 'a few times a week' and 'daily' (5).

For this study questions about social desirability were added to the questionnaire of the RYM. Social desirability was measured by a subscale of the School Questionnaire, consisting of eight items; for example 'I sometimes tell a lie', 'I like all my teachers', 'I'm always polite to adults' (Smits and Vorst, 1982). Answering categories were: 'that's true' (0), 'I don't know' (1), 'that's not true' (2). Of these eight items a sum score was calculated, ranging from 0 to 16. Cronbach's α in this sample was 0.59. To include social desirability as a group factor in the analysis, the score was dichotomised (80th percentile (scores above 8) versus lower scores).

Data analysis

To study the differences in responses of well-being, behaviour and related factors (stressful life events, school factors) between the anonymous and confidential condition, ANOVA's (One-way Analysis of Variance) were performed. Dependent variables were the indicators of well-being, behaviour and the related factors. Gender (boys, girls), condition (anonymous, confidential), social desirability (high, neutral and low) and the interaction effects of 'condition and gender' and 'condition and social desirability' were entered in the analysis as independent variables. Since level of education and age were not distributed equally between the two conditions, we added these in the analysis as independent variables to control for any potential confounding effect. For this purpose age was dichotomised in '16 years or younger' and '17 years or older'. The strength of the difference in responses between the method of data collection was expressed by computing 'effect sizes' considering f=0.10 as a small, f=0.25 as a medium, and f=0.40 as a large difference (Cohen, 1988).

Results

For each variable means and standard deviations were computed by gender and condition (Table 2.2). Results from the analyses of variances are also presented in table 2.2. The results presented are limited to the main effect of condition, the interaction effect of condition and gender and the interaction effect of condition and social desirability. All analyses are controlled for level of education and age. Analysis of the indicators of well-being showed a significant main effect for perceived health: pupils in the confidential group perceived their health (in general) as less positive than pupils in the anonymous group. The effect size of this difference did reach Cohen's level of 'small'. For all other indicators of well-being no significant differences between the groups were found.

For the problem behaviour items significant condition-by-gender interactions were found for aggressive behaviour, vandalism and stealing and violent delinquent behaviour. Boys in the confidential group reported significant less aggressive behaviour, vandalism and stealing and violent delinquent behaviour than boys in the anonymous group. For girls no such differences occurred. However, only for aggressive behaviour the effect size was small. For the other problem behaviours (vandalism, stealing, violent delinquent behaviour) the effect sizes did not reach above 0.10.

On the health risk behaviours a main effect was found for use of alcohol: pupils in the confidential condition reported significantly less use of alcohol than pupils in the anonymous condition. Also, a significant interaction effect of condition and social desirability was found: pupils with high scores on social desirability reported less use of alcohol in the confidential condition than in the anonymous condition. The effect size of these differences did reach Cohen's level of 'small'.

Of the related factors a significant interaction effect of condition and gender was found for 'frequent fights parents'. Girls in the confidential group reported less frequent fights between their parents than girls in the anonymous group (effect size less than small). For boys no difference between the groups was found. For physical abuse, a significant interaction effect of condition and social desirability was found; pupils with high scores on social desirability reported more physical abuse in the confidential condition than in the anonymous condition. 'Victim offence' was reported less frequently in the confidential condition than in the anonymous condition (effect size less than small). Further, boys perceived the relationship with teachers more positively in the confidential condition, whereas girls were more positively in the anonymous condition.

Table 2.2 Means (SD) by gender and condition; results ANOVA: main effect condition, interaction effect condition and gender

	Boys			
	Anon n=165		Conf n=165	
Measure	*Mean*	*(SD)*	*Mean*	*(SD)*
Well-being				
self-esteem (0-10)	8.8	(1.9)	9.0	(1.6)
suicidal ideation (1-5), past year	1.4	(0.9)	1.2	(0.8)
suicide attempt (1-3), ever	1.1	(0.4)	1.0	(0.3)
mental health (0-12)	1.6	(2.4)	1.8	(2.6)
perceived health (1-5)	1.7	(0.7)	1.7	(0.7)
Problem behaviour				
aggressive behaviour (0-16)	2.3	(3.5)	1.3	(2.3)
vandalism and stealing (0-24)	2.4	(4.9)	1.5	(2.6)
violent delinquent behaviour (0-16)	0.8	(2.8)	0.3	(1.3)
carrying a weapon (1-3)	1.3	(0.6)	1.2	(0.5)
playing truant (1-6)	2.3	(1.4)	1.9	(1.2)
bullying (1-5)	1.7	(1.1)	1.6	(0.9)
Health risk behaviour				
smoking (1-5), past 4 weeks	2.4	(1.5)	2.2	(1.4)
use of alcohol (1-6), past 4 weeks	3.2	(1.7)	2.9	(1.5)
use of cannabis (1-6), past 4 weeks	1.7	(1.4)	1.5	(1.3)
use of hard drug (0-20), past 4 weeks	0.8	(3.2)	0.3	(2.3)
Stressful life events and school factors				
death family member (1-2)	1.1	(0.3)	1.1	(0.3)
death significant other (1-2)	1.7	(0.5)	1.6	(0.5)
severe illness – self (1-2)	1.1	(0.3)	1.1	(0.3)
severe illness – father/mother (1-2)	1.2	(0.4)	1.2	(0.4)
psychological problems parent (1-2)	1.1	(0.3)	1.1	(0.3)
frequent fights between parents (1-2)	1.3	(0.5)	1.3	(0.5)
divorce parents (1-2)	1.2	(0.4)	1.1	(0.3)
unwanted pregnancy (1-2)	1.0	(0.2)	1.0	(0.1)
physical abuse (1-2)	1.1	(0.3)	1.1	(0.3)
sexual abuse (1-2)	1.0	(0.2)	1.0	(0.1)
victim offence (1-2)	1.4	(0.5)	1.3	(0.5)
problems at home (0-5)	0.4	(0.9)	0.5	(0.9)
perception of academic functioning (0-100)	71.8	(18.8)	74.0	(17.8)
perception of relationship with classmates (0-100)	85.8	(14.8)	88.2	(11.8)
perception of relationship with teachers (0-100)	66.8	(18.2)	71.2	(17.1)
victim of bullying at school (1-5)	1.3	(0.8)	1.2	(0.5)

Note: The answering scale-ranges for each of the variables are presented in parentheses after the variable name.

Girls				Ana lysis of Variance					
Anon n=186		Conf n=188		Condition		Condition x gender		Condition x social desirability	
Mean	*(SD)*	*Mean*	*(SD)*	*F*	*f*	*F*	*f*	*F*	*f*
7.9	(2.6)	8.0	(2.5)	2.96		0.07		2.95	
1.3	(0.7)	1.3	(0.6)	0.06		0.30		0.63	
1.1	(0.3)	1.0	(0.2)	0.68		0.01		1.83	
3.0	(3.0)	2.8	(3.3)	1.46		0.67		0.40	
2.0	(0.7)	2.1	(0.8)	6.81**	0.10	0.18		1.83	
0.3	(0.9)	0.3	(0.9)	0.36		9.68**	0.12	0.43	
0.4	(1.1)	0.4	(1.0)	0.36		4.32*	0.08	0.97	
0.0	(0.3)	0.0	(0.1)	2.77		5.70*	0.09	0.29	
1.1	(0.2)	1.0	(0.2)	0.06		0.22		0.02	
1.9	(1.1)	1.7	(1.1)	0.05		0.70		0.78	
1.3	(0.6)	1.2	(0.4)	0.18		0.02		1.23	
2.5	(1.5)	2.1	(1.3)	0.72		0.40		0.53	
2.4	(1.2)	2.3	(1.2)	6.89**	0.10	1.73		9.62**	0.12
1.3	(0.9)	1.1	(0.5)	0.46		0.02		0.56	
0.0	(0.2)	0.0	(0.2)	1.93		1.85		0.41	
1.1	(0.2)	1.1	(0.3)	3.77		0.25		1.41	
1.6	(0.5)	1.6	(0.5)	0.24		0.11		2.10	
1.1	(0.4)	1.1	(0.3)	0.96		0.29		0.14	
1.2	(0.4)	1.2	(0.4)	1.01		0.03		2.04	
1.1	(0.4)	1.1	(0.3)	1.63		0.46		0.04	
1.4	(0.5)	1.3	(0.5)	0.01		4.25*	0.08	0.00	
1.2	(0.4)	1.1	(0.4)	2.98		0.01		1.40	
1.0	(0.2)	1.0	(0.1)	0.03		0.12		0.49	
1.1	(0.3)	1.1	(0.2)	0.68		0.28		5.87*	0.09
1.1	(0.3)	1.1	(0.3)	0.04		0.02		0.09	
1.2	(0.4)	1.1	(0.3)	3.91*	0.08	0.75		0.99	
0.7	(1.2)	0.6	(1.1)	0.76		1.12		1.84	
70.4	(19.4)	73.6	(18.3)	0.89		0.10		2.30	
86.5	(10.9)	87.5	(11.5)	1.13		0.71		0.58	
73.3	(13.7)	73.5	(13.8)	0.30		3.97*	0.08	0.08	
1.2	(0.5)	1.2	(0.5)	3.53		0.21		0.59	

** p<0.05 ** p<0.01*

Discussion

This study compared the responses of mainly 16-17 year old adolescents on self-report measures of various (psychological) health indicators in two conditions: anonymous and confidential collection. Both conditions differed in the verbal instructions given before the administration in the classroom and the questionnaire itself, names being printed on the confidential version. We found that the differences in response between the two groups were small for most items and scales. Out of 93 significance tests, four emerged as significant at $p<0.01$, and a further six at $p<0.05$. Effect sizes corresponding with six of these significant tests remained minimal. Thus only the main effect of perceived health, use of alcohol, condition-by-gender interaction of aggressive behaviour and condition-by-social desirability interaction of use of alcohol showed an effect size that reached Cohen's (1988) level of 'small'. Hence we conclude that in general, confidential administration of these self-report questionnaires renders the same epidemiological results as anonymous administration.

If our conclusion were completely valid, the implications would be substantial. Nevertheless, there were significant differences between the two groups for several items – namely, perceived health, aggressive behaviour, vandalism and stealing, violent delinquent behaviour, use of alcohol, the report of frequent fights between parents, being a victim of an offence, physical abuse and perception of relationship with teachers. It seems that self-report of the more externalising problems, especially for boys, is likely to differ from confidential report, confirming the results of Kulik et al. (1968). However, these differences did not exist for similar (sensitive) variables, such as carrying a weapon, use of hard drugs, bullying and truancy. Furthermore, the effect sizes remained minimal, which implies that only a minor proportion of the variance can be explained by the experimental condition.

For boys as well as girls, this study showed no difference between anonymous or confidential data collection regarding smoking, use of marihuana and use of hard drug. This is in accordance with some previous studies on substance use (King, 1970; O'Malley et al., 2000; Ong and Weiss, 2000). We don't have an explanation for the difference in use of alcohol which is reported more frequently in the anonymous condition. This finding doesn't confirm the results of O'Malley et al. (2000) and Malvin and Moskowitz (1983). Perhaps the use of alcohol is a more sensitive subject than we supposed.

Further, the results indicate that the effect of social desirability is not different in the anonymous and confidential condition, which confirmed the results of Hancock and Flowers (Hancock and Flowers, 2001). This means that pupils scoring high on social desirability do not report different at most health indicators in the two conditions.

A few remarks should be made about the experimental condition itself. The design of the study compared an anonymous with a confidential administration of self-report questionnaires. A limitation of the generalisability of the present study may result from our operationalisation of confidentiality and anonymity. In this study, perhaps more than in other studies, a lot of effort has been put into demonstrating the confidentiality of the questionnaires: using

a blank sealed envelope and a thorough verbal instruction. During the study and previous assessments of the Rotterdam Youth Monitor, remarks by some pupils were made which revealed suspicion about the anonymity in the anonymous condition, despite the assurances. A further (follow-up) validation of these results seems interesting in this respect. Another remark concerns the age of the subjects. Although we think of these older adolescents as being the most critical (and suspicious) among our entire population at secondary education schools, we do not have proof of that at this moment. Further, similar studies in lower grades and other types of education should confirm the results.

Despite these reservations, this research shows that for most health characteristics, epidemiologically useful outcomes may be derived from confidential self-report measures. The lack of findings on our measure for social desirability supports this conclusion, and hence it is safe to use questionnaires to detect individual at-risk children as well as to monitor youth's health in general. Given that a recently introduced Dutch law has stated that government(s) should monitor the health of young people, both individually and collectively, a confidential administration of a self-report may be a feasible method of compliance. Finally, externalising problem behaviours, which seem to be underrated by the confidential method, are unlikely to escape attention in schools. A recommendation would be to make sure that the confidentiality of the questionnaires should be guaranteed and communicated accordingly. Separate tables, using bar codes instead of names, using envelopes to put the questionnaires in after completion and clear statements about the confidentiality during the instruction may all help to give the most accurate results.

References

Bjarnason T, Adalbjarnardottir S (2000), Anonymity and confidentiality in school surveys on alcohol, tobacco and cannabis use. *J Drug Issues* 30: 335-344.

CBS (2001), *Jeugd, cijfers en feiten* Voorburg/Heerlen: CBS.

Cohen J (1988), *Statistical power analysis for the behavior sciences (Rev. edn.)* New York: Academic.

De Zwart R (1999), *Nationaal scholierenonderzoek 1996. Deel 2: Gedrag en Gezondheid van scholieren.* Utrecht: NIBUD.

Garnefski N, Diekstra RFW (1993), *Scholierenonderzoek 1992: landelijke uitkomsten rond een onderzoek naar gedrag en gezondheid van ca. 11.000 scholieren van het voortgezet onderwijs.* Leiden: Rijksuniversiteit Leiden.

Hancock DR, Flowers CP (2001), Comparing social desirability responding on world wide web and paper-administered surveys. *Educ Technol Res Dev* 49: 5-14.

Junger-Tas J, van der Laan PH, Kruisink M (1992), *Ontwikkeling van de jeugdcriminaliteit en de justitiële jeugdbescherming: periode 1980-1990* Arnhem, The Netherlands: Gouda Quint.

King FW (1970), Anonymous versus identifiable questionnaires in drug usage surveys. *Am Psychol* 25: 982-985.

Koeter MWJ, Ormel J (1991), *General Health Questionnaire. Dutch version, manual* Lisse: Swets and Zeitlinger.

Kulik JA, Stein KB, Sarbin TR (1968), Disclosure of delinquent behavior under conditions of anonymity and nonanonymity. *J Consult Clin Psychol* 32: 506-509.

Malvin JH, Moskowitz JM (1983), Anonymous versus identifiable self-reports of adolescent drugs attitudes, intentions, and use. *Public Opin Q* 47: 557-566.

O'Malley PM, Johnston LD, Bachman JG, Schulenberg J (2000), A comparison of confidential versus anonymous survey procedures: Effects on reporting of drug use and related attitudes and beliefs in a national study of students. *J Drug Issues* 30: 35-54.

Ong AD, Weiss DJ (2000), The impact of anonymity on responses to sensitive questions. *J Appl Psychol* 30: 1691-1708.

Rosenberg M (1965), *Society and the Adolescent Self-image.* Princeton: Princeton University Press.

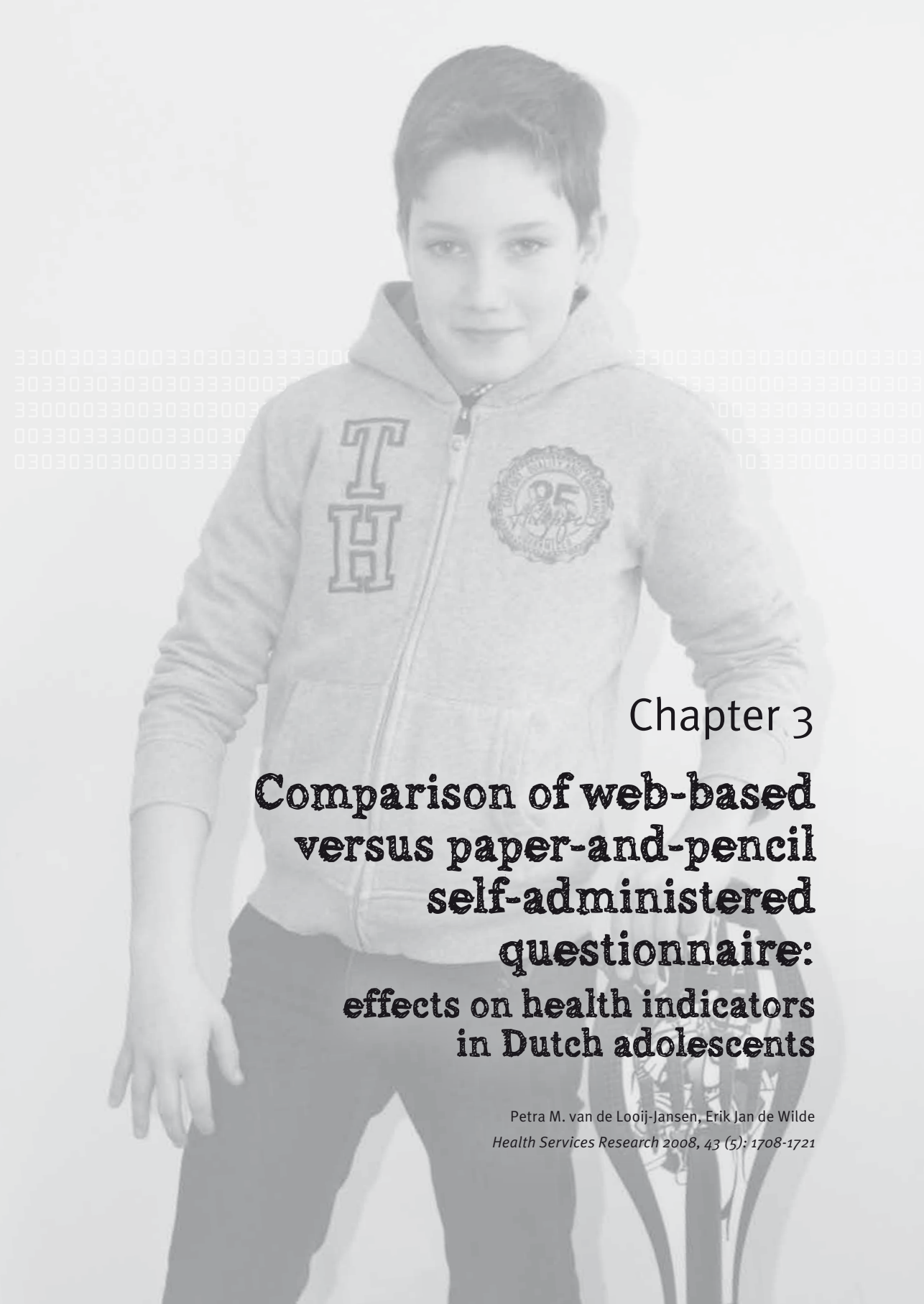
Smits JAE, Vorst HCM (1982), *Schoolvragenlijst voor basisonderwijs en voortgezet onderwijs. Handleiding voor gebruikers* Nijmegen: Berkhout B.V.

Stoel WGR (1980), *De relatie tussen de grootte van scholen voor voortgezet onderwijs en het welbevinden van de leerlingen. Deel I: Resultaten van een empirisch onderzoek in de Nederlandse onderwijssituatie. Deel II: De ontwikkeling van een schoolbelevingsschaal.* Haren: Research Instituut voor het Onderwijs in het Noorden (RION).

Tourangeau R, Rasinski K, Jobe JB, Smith TW, Pratt W (1997), Sources of error in a survey of sexual behavior. *J Off Stat* 131: 341-365.

Tourangeau R, Rips L, Rasinski K (2000), *The psychology of survey response.* Cambridge, UK: Cambridge University Press.

Van der Linden FJ, Dijkman, TA, and Roeder, PJB. (1983), *Schoolgaande jongeren, hun leefwereld en zelfbeleving.* Nijmegen: Hoogveldt Instituut.



Chapter 3

Comparison of web-based versus paper-and-pencil self-administered questionnaire:

effects on health indicators in Dutch adolescents

Petra M. van de Looij-Jansen, Erik Jan de Wilde

Health Services Research 2008, 43 (5): 1708-1721

Abstract

***Background*: To investigate differences in responses related to (mental) health and health-related behaviour between two methods of data collection: web-based (web) and paper-and-pencil (p&p). Privacy and confidentiality concerns were investigated as a possible source of response bias.**

***Methods*: The study was conducted as part of the Rotterdam Youth Monitor, a longitudinal health surveillance system. Within each participating school all third-grade classes (mainly 14-15 year olds pupils) were randomly assigned to either the web-based condition (n=271) or the paper-and-pencil condition (n=261). Pupils completed the confidential questionnaire at school in a computer lab or classroom. Response rate was 90%.**

***Results*: For most indicators of well-being and behaviour no differences between the two conditions were found, except for SDQ emotional symptoms (p&p>web), SDQ prosocial behaviour (p&p>web) and carrying a weapon (web>p&p). Although significant, the actual size of these differences just reached the criteria of Cohen for a small (or less than small) difference. Perceived level of privacy and confidentiality did not differ between the two modes.**

***Conclusions*: The findings suggest that in a controlled school setting, web-based administration of health indicators yields almost the same results as paper-and-pencil administration. To generalise these findings, we recommend repeated studies in other populations and settings.**

Introduction

Due to the many advantages, computerized questionnaires are used more and more in youth surveys, replacing the more traditional paper-and-pencil questionnaires. Advantages of computerized questionnaires are, for instance, the possibility to check directly for missing answers and consistency, automatic branching, cost advantages, and the possibility to give computerized tailored feedback on health and health-related behaviour, which may be more effective than generic (paper) health advices (Brug et al., 1999). As the method of data collection can affect the answers that are obtained, especially for sensitive questions (Bowling, 2005; Tourangeau and Smith, 1996), it is important to determine whether responses to computerized questionnaires are comparable to those obtained by paper-and-pencil method.

So far, findings from previous studies among school-aged children and adolescents are inconclusive. Several studies did not find major differences between computerized and paper-and-pencil questionnaires (Hays and McCallum, 2005; Mangunkusumo et al., 2005; Mangunkusumo et al., 2006; McCabe et al., 2005; Truman et al., 2003), whereas other studies showed that adolescents disclose more sensitive information in computerized questionnaires than in paper-and-pencil conditions (Paperny et al., 1990; Turner et al., 1998; Wang et al., 2005; Wright et al., 1998) or found (gender-related) differences for some topics (Beebe et al., 1998; Hallfors et al., 2000; Vereecken and Maes, 2006; Webb et al., 1999). Of these studies a few were web-based (Mangunkusumo et al., 2005; Mangunkusumo et al., 2006; McCabe et al., 2005; Wang et al., 2005).

Several factors may be responsible for the differences found in these studies. It is known that the validity of self-reports may be affected by cognitive as well as situational factors. Factors considered especially influential include the presence of others while responding to questions and the respondents' perceptions of privacy or confidentiality (Bowling, 2005; Brener et al., 2003). A perceived lack of privacy or confidentiality could cause response bias because of a fear of reprisal. In particular, behaviour that is illegal, stigmatized, or laden with moral implications may be underreported because of this concern (Brener et al., 2003).

As the setting of computerized administration of questionnaires at schools is likely to be different from the administration of paper-and-pencil questionnaires, it is important to investigate respondents' perceived level of privacy and confidentiality as a possible source of response bias. However, little is known about adolescents' perceived level of privacy and confidentiality in surveys. Because of this and the inconsistent pattern which appears from the literature, the following issues were examined in the present study: a) (gender-related) differences in perceived level of privacy and confidentiality between web-based and paper-and-pencil questionnaires; and b) (gender-related) differences in the report of indicators of (mental) health and behaviour between web-based and paper-and-pencil questionnaires when taking into account pupils' perceived level of privacy and confidentiality.

Methods

Study design and procedure

The present study was conducted as part of the Rotterdam Youth Monitor (RYM), a longitudinal youth health surveillance system. The RYM monitors the general health, well-being, behaviour, and related factors of youth aged 0 to 19 years living in Rotterdam and surroundings (the Netherlands) in order to supply information for youth policy at the school, neighborhood and municipality level. The RYM is incorporated in the regular check-ups of the preventive youth healthcare system.

Five secondary schools with various educational levels were approached (and all agreed) for participation in this study. Within each school all third-grade classes (n=26) were randomly assigned to either the internet condition or the paper-and-pencil condition (p&p). Pupils assigned to the internet condition completed the questionnaire in the school computer lab.

Four specially trained school nurses of the Municipal Public Health Service carried out the administration of the RYM questionnaire during regular class hours. Pupils were assured that all information provided was strictly confidential. It was stated (verbal and on paper) that completed questionnaires would "never be shown to parents, teachers, police or anyone else, except to the school nurse". Pupils were also told they could be invited for a health examination by the school nurse later on in the school year. After this introduction, pupils in the paper-and-pencil condition were given a questionnaire with their name printed on the front page. Pupils in the internet condition were given an instruction sheet with their name, a username, a login code and the name of the website (*www.jeugdmonitorrotterdam.nl*).

The two versions of the questionnaire were identical in terms of the questions asked, their wording, and their order of presentation in the survey. The computerized version differed from the paper-and-pencil version in the way the answering categories were presented. In the computerized version all answering categories were presented below each question, whereas in the paper-and-pencil version the answering categories of some questions were presented next to the question (in order to save printing space). Per screen multiple questions of the same topic were presented. Questions that were not relevant to the pupil were not displayed. Logging out after completing the questionnaire was allowed only after answering all items.

Parents received an information letter and were given the opportunity to refuse their child's participation. In the Netherlands, for this kind of research (as part of a routine health examination offered to all children) informed consent is legally required.

The questionnaire was filled out in the classroom in about one hour in the presence of a teacher (for keeping order in the classroom) and a school nurse. A researcher was present when the internet version was administered (for technical problems). Data were collected in November and December 2005.

Table 3.1 Socio-demographic characteristics of the study population, by mode of administration

	internet		paper-and-pencil			
	n	%	n	%	χ^2 (df)	p
Gender						
boys	152	56.3%	127	48.7%	3.10 (1)	0.08
girls	118	43.7%	134	51.3%		
Age						
14 years and younger[1]	150	55.6%	133	51.2%	1.17 (2)	0.56
15 years	100	37.0%	108	41.5%		
16 years and older	20	7.4%	19	7.3%		
Level of education						
basic pre-vocational training	119	44.1%	118	45.2%	0.15 (3)	0.99
theoretical pre-vocational training	20	7.4%	20	7.7%		
general secondary education	75	27.8%	72	27.6%		
pre-university education	56	20.7%	51	19.5%		
Ethnic background						
Dutch	113	41.9%	121	46.7%	3.03 (4)	0.55
Surinamese/Dutch Antillean/Aruban	39	14.4%	33	12.7%		
Moroccan	33	12.2%	28	10.8%		
Turkish	36	13.3%	25	9.7%		
Other	49	18.2%	52	20.1%		

[1] *5 pupils were 13 years old*

Subjects

Overall response rate was 90% and did not differ by administration mode. Reasons for absence were mainly illness. A total of 532 pupils completed the questionnaire, 271 in the internet condition and 261 in the paper-and- pencil condition. The data of 531 cases were analysed, because the data of one pupil (internet condition) were not reliable. Socio-demographic characteristics of the sample were equally distributed across internet mode and paper-and-pencil mode ($p \geq 0.05$, Table 3.1).

Measures

(Mental) health and problem behaviour

The Dutch self-report version of the Strengths and Difficulties Questionnaire (SDQ) was used (van Widenfelt et al., 2003). The SDQ comprises of five subscales of five items each: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour. Self-esteem was measured by the Dutch version of the Rosenberg Self-esteem Scale (Rosenberg, 1965; van der Linden, 1983). Psychological well-being was measured with nine items about feelings and moods, derived from the Child Health Questionnaire (CHQ-CF87). (Landgraf et al., 1996; Raat et al., 2002). Perceived health was assessed by means of one question in which the respondents were asked to rate their health (CBS, 2001). Suicidal ideation was assessed with one item: 'In the last 12 months, have you thought about putting an end to your life?' (Garnefski and Diekstra, 1993). Suicide attempt was measured by the question: "Did you ever seriously attempt putting an end to your life?".

Furthermore, the questionnaire included four items about aggressive behaviour and six items about delinquent behaviour, developed and tested by the scientific research department of the Ministry of Justice (Junger-Tas et al., 1992), one question about carrying a weapon, one question about playing truant and one question about bullying.

Health risk behaviour and sexual behaviour

Questions on smoking, alcohol consumption and cannabis use were based on a questionnaire used by the Netherlands Institute of Mental Health and Addiction (De Zwart, 1999). Smoking habits were measured by the question: "Have you ever smoked cigarettes?". For alcohol consumption we used the question concerning frequency of consumption in the past four weeks. Cannabis use was measured by the question "How often have you used cannabis in the past four weeks". Sexual behaviour was measured by the question: "Have you ever had sexual intercourse?". This question was not further specified.

Perceived level of privacy and confidentiality

Perceived level of privacy and confidentiality was measured by two questions: 'I had sufficient privacy when completing this questionnaire' and 'I trust that the Municipal Health Service will treat my data confidentially' (1=fully agree, 4=fully disagree). A sum score of these two questions was calculated (Cronbach's alpha 0.66).

Data analysis

Differences in responses on the health indicators between the two modes of administration were analysed by ANCOVA's (two-way Analysis of Covariance). Condition (internet, paper-and-pencil), gender (boys, girls) and the interaction between condition and gender were entered in the analysis as independent variables. Perceived level of privacy and confidentiality was

entered into the analyses as covariate in order to control for this factor as a source of bias. Before carrying out these analyses we tested one of the assumptions for ANCOVA's, e.g. that the relationship between perceived level of privacy and confidentiality and the dependent variables is the same across the two modes of administration. These results showed no interaction effect between perceived level of privacy and confidentiality and the health indicators. Missing data were less than 5% and were excluded. Effect sizes were computed, considering f=0.10 as a small, f=0.25 as a medium, and f=0.40 as a large difference (Cohen, 1988).

Results

No significant differences in perceived level of privacy and confidentiality between web-based and paper-and-pencil questionnaires were found (Table 3.2) and this did not differ by gender (data not shown).

Table 3.2 Perceived level of privacy and confidentiality, by mode of administration

		internet n=270	paper-and-pencil n=261	p-value[1]
I had sufficient privacy when completing this questionnaire	fully agree	30.7%	33.7%	0.56
	agree	51.9%	49.4%	
	disagree	14.1%	12.9%	
	fully disagree	3.3%	4.0%	
I trust that the Municipal Health Service will treat my data confidentially	fully agree	47.4%	43.4 %	0.26
	agree	44.1%	45.0%	
	disagree	4.8%	8.0%	
	fully disagree	3.7%	3.6%	

[1] *Mann-Whitney U-test*

Table 3.3 Means (SD) by gender and condition; results ANCOVA: main effect condition, interaction effect condition and gender, main effect confidentiality

Measure	Boys			
	Internet n=152		P&P n=127	
	Mean	*(SD)*	*Mean*	*(SD)*
(Mental) health and behaviour				
Strengths and Difficulties Questionnaire				
total difficulties (sum score 0-40)[a]	8.7	(3.9)	9.7	(5.2)
emotional symptoms (sum score 0-10)	1.6	(1.4)	2.1	(1.9)
conduct problems (sum score 0-10)	2.0	(1.3)	2.0	(1.4)
hyperactivity-inattention (sum score0-10)	3.6	(2.3)	3.8	(2.5)
peer problems (sum score 0-10)	1.5	(1.3)	1.8	(1.6)
prosocial behaviour (sum score 0-10)	6.3	(1.9)	6.7	(1.9)
self-esteem (sum score 0-10)	8.7	(1.6)	8.4	(1.9)
psychological well-being CHQ (sum score 0-100)	80.5	(10.3)	81.0	(12.4)
perceived health (1-5)	1.8	(0.6)	1.8	(0.8)
suicidal ideation (1-5), past year	1.1	(0.4)	1.2	(0.5)
suicide attempt (1-3), ever	1.0	(0.1)	1.1	(0.3)
aggressive behaviour (sum score 0-16), past year	1.5	(2.3)	1.5	(2.5)
vandalism and stealing (sum score 0-24), past year	1.3	(2.1)	1.1	(2.2)
carrying a weapon (1-3)	1.1	(0.4)	1.1	(0.3)
playing truant (1-6), past 4 weeks	1.2	(0.6)	1.1	(0.4)
bullying (1-5), past 4 weeks	1.3	(0.6)	1.5	(0.8)
Health risk behaviour				
smoking (1-5)	1.6	(0.9)	1.6	(1.1)
use of alcohol (1-6), past 4 weeks	1.6	(0.9)	1.6	(1.0)
use of marijuana (1-6), past 4 weeks	1.1	(0.4)	1.2	(0.6)
Sexual behaviour (1-4)	1.5	(1.0)	1.3	(0.8)

** p ‹ 0.05, ** p ‹ 0.01, *** p ‹ 0.001; [1] p&p › internet, [2] internet › p&p, [3] high score ‹ privacy and confidentiality, [4] high score › privacy and confidentiality*

Note: The scale range for answers to each of the variables/sum scores are presented in parentheses after the variable name.

[a] *A total difficulties score on the SDQ is derived by summing four subscales relating to emotional symptoms, conduct problems, hyperactivity-inattention, and peer problems.*

	Girls				Analysis of Variance					
	Internet n=118		P&P n=134		Condition		Condition x gender		Perceived privacy and confidentiality	
	Mean	*(SD)*	*Mean*	*(SD)*	*F*	*f*	*F*	*f*	*F*	*f*
	10.3	(5.0)	10.0	(5.0)	0.60		1.32		4.78*[3]	0.10
	3.0	(2.1)	3.3	(2.2)	3.92*[1]	0.09	0.07		0.00	
	1.6	(1.4)	1.6	(1.3)	0.03		0.04		0.77	
	3.8	(2.6)	3.5	(2.3)	0.30		1.02		9.40**[3]	0.14
	1.8	(1.6)	1.7	(1.5)	0.86		2.15		1.47	
	7.8	(1.7)	8.2	(1.5)	8.56**[1]	0.13	0.20		13.3***[4]	0.16
	7.8	(2.6)	8.0	(2.3)	0.00		1.68		2.33	
	73.0	(16.6)	75.0	(15.9)	1.53		0.16		0.14	
	2.0	(0.7)	2.0	(0.7)	0.01		0.08		0.06	
	1.5	(0.9)	1.4	(0.8)	0.51		1.64		0.03	
	1.1	(0.4)	1.1	(0.4)	0.02		0.40		0.35	
	0.6	(1.4)	0.5	(1.1)	0.75		0.01		2.15	
	0.7	(1.8)	0.4	(1.0)	2.75		0.01		2.71	
	1.1	(0.3)	1.0	(0.1)	4.87*[2]	0.10	0.03		0.69	
	1.1	(0.5)	1.2	(0.5)	0.01		3.56		0.13	
	1.1	(0.5)	1.2	(0.5)	3.46		0.98		4.82*[3]	0.10
	1.7	(1.1)	1.6	(1.0)	0.38		0.72		0.00	
	1.6	(0.9)	1.4	(0.8)	0.62		1.30		0.36	
	1.1	(0.6)	1.1	(0.4)	0.03		0.34		0.01	
	1.2	(0.7)	1.2	(0.6)	1.61		0.42		1.23	

Table 3.3 shows the means (SD) of the health indicators, by mode of administration and gender, and the results of the analyses of covariance. For two subscales of the SDQ differences were found. SDQ emotional symptoms were reported less frequently in the internet condition. Pupils in the paper-and-pencil condition reported more prosocial behaviour compared to pupils in the internet condition. For both differences effect sizes reached Cohen's level of 'small'. For all indicators of well-being no differences were found. Of the problem behaviour items, main effects of condition were found for 'carrying a weapon'. Pupils in the internet condition reported significantly more often carrying a weapon compared to pupils in the paper-and-pencil questionnaire (small effect size). No differences were found for smoking, use of alcohol, use of marijuana and sexual behaviour.

Finally, as can be seen from Table 3.3, a high score on SDQ total difficulties, SDQ hyperactivity-inattention and bullying (indicating more problems) was associated with less privacy and confidentiality. A high score on SDQ prosocial behaviour was associated with more perceived privacy and confidentiality.

Discussion

For most indicators of (mental) health and behaviour no significant (gender-related) differences between the two modes were found, which is in agreement with previous research reporting no or very few differences between computerized and paper-and-pencil studies (Hays and McCallum, 2005; Mangunkusumo et al., 2005; Mangunkusumo et al., 2006; McCabe et al., 2005; Truman et al., 2003). Significant but small differences between the two modes of data collection were found for the SDQ subscales 'emotional symptoms' (p&p>web) and 'prosocial behaviour' (p&p>web), and 'carrying a weapon' (web>p&p). Furthermore, no difference in perceived level of privacy and confidentiality between the two conditions was found.

Our results do not confirm the finding from other studies (Paperny et al., 1990; Turner et al., 1998; Wang et al., 2005; Wright et al., 1998) that adolescents disclose more sensitive information in computerized questionnaires. For sexual behaviour, which can be regarded as the most sensitive topic in our survey, no differences between web-based and paper-and-pencil mode were found. Also for other sensitive topics like use of alcohol, use of marijuana, vandalism and stealing no differences were found. The only exception is found for carrying a weapon, which is reported more often in the web-based condition.

Although no differences in perceived level of privacy and confidentiality between the two modes were found, the percentage of pupils who do not agree with the statement 'I had sufficient privacy when completing this questionnaire', is high, i.e. about 17% in both conditions. For confidentiality assurances this percentage is somewhat lower (about 10%), but still relatively high, despite the efforts we made to assure the confidentiality of the survey, e.g. training school nurses for the administration, closed boxes to put the questionnaires in after the paper-and-pencil administration. Given these results it seems likely that some pupils were

not sufficiently convinced of the confidentiality of the study. In addition, the knowledge they could be invited for a health examination might have influenced their answer on the confidentiality question. As no comparable data are available, it is difficult to interpret these findings. Furthermore, the association of perceived level of privacy and confidentiality and the reporting of (sensitive) topics appeared to be limited, except for a few topics. Moreover, the association was not in the direction we expected. For example, a high SDQ total difficulties score was associated with less perceived privacy and confidentiality. This may be due to the order of the questions (Dilman, 2007; Serdula et al., 1995), because questions about privacy and confidentiality were asked at the end of the questionnaire. Perhaps these questions reflect pupils' need of privacy and confidentiality more than their actual sense of confidentiality. Further research is needed to investigate which factors play a role in pupils' concern regarding privacy and confidentiality.

A few limitations should be mentioned. The findings are applicable to the school setting as the use of web-based versus paper-and-pencil questionnaires in clinical settings and in other age groups was not studied. To generalize these findings, we recommend repeated studies in other populations and settings.

Concerning the study design, classrooms were randomly allocated to one of the two conditions. A disadvantage of this design is that within-pupils' variance was not taken into account. A randomized crossover design would be more applicable. However, crossover designs may induce a carry-over effect, i.e. administration during the first period may carry over into the second administration period.

In the Netherlands one of the main legal tasks of preventive youth healthcare is to monitor youth health and detection of health risks. In Rotterdam this is performed both individually and collectively. For third-grade pupils the traditional paper-and-pencil questionnaire of the RYM (SDQ total score) is used as a first part of a two-step screening in order to pre-select pupils for a visit to the school nurse. In the second step, the school nurse assesses the need for further referral, often with clinical screening measures. This two-step screening procedure is used because it is known that several clinical screening measures profit significantly in positive predictive power when used in groups that were pre-selected by a general screening measure (Loeber et al., 1984; Rothman and Greenland, 1998). Our results show that this general screening procedure, as the SDQ total difficulties is concerned, is also possible by using computerized questionnaires in a school setting.

In conclusion, our study provides further evidence that in a controlled school setting, web-based administration of health indicators yields almost the same results as paper-and-pencil administration. However, this study also demonstrates that in general (regardless the method of data collection) pupils' sense of privacy and confidentiality is associated with some of their answers. Therefore, researchers should pay attention to guarantee sufficient privacy and confidentiality in classroom settings.

Acknowledgements

The authors thank Joop Duisterhout, MSc (Erasmus Medical Centre, Department of Medical Informatics) for adjusting the Internet tool, and Resi Mangunkusumo, PhD (Erasmus Medical Centre, Department of Public Health) for her advice and help. We also thank the school nurses of the Municipal Public Health Service for administering the questionnaires and the participating pupils for filling out the questionnaire.

References

Beebe TJ, Harrison PA, McRae JA, Jr., Anderson RE, Fulkerson JA (1998), An evaluation of computer-assisted self-interviews in a school setting. *Public Opin Q* 62: 623-632.

Bowling A (2005), Mode of questionnaire administration can have serious effects on data quality. *J Publ Health (Oxf)* 27: 281-91.

Brener ND, Billy JOG, Grady WR (2003), Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *J Adolesc Health* 33: 436-457.

Brug J, Campbell M, van Assema P (1999), The application and impact of computer-generated personalized nutrition education: A review of the literature. *Patient Educ Couns* 36: 145-156.

CBS (2001), *Jeugd, cijfers en feiten* Voorburg/Heerlen: CBS.

Cohen J (1988), *Statistical power analysis for the behavior sciences (Rev. edn.)* New York: Academic.

De Zwart R (1999), *Nationaal scholierenonderzoek 1996. Deel 2: Gedrag en Gezondheid van scholieren.* Utrecht: NIBUD.

Dilman DA (2007), *Mail and internet surveys. The tailored design method (second edition)* Hoboken, New Jersey: John Wiley & Sons, Inc.

Garnefski N, Diekstra RFW (1993), *Scholierenonderzoek 1992: landelijke uitkomsten rond een onderzoek naar gedrag en gezondheid van ca. 11.000 scholieren van het voortgezet onderwijs.* Leiden: Rijksuniversiteit Leiden.

Hallfors D, Khatapoush S, Kadushin C, Watson K, Saxe L (2000), A comparison of paper vs computer-assisted self interview for school alcohol, tobacco, and other drug surveys. *Eval Program Plann.* 23: 149-155.

Hays S, McCallum RS (2005), A Comparison of the Pencil-and-Paper and Computer-Administered Minnesota Multiphasic Personality Inventory-Adolescent. *Psychol Schools* 42: 605-613.

Junger-Tas J, van der Laan PH, Kruisink M (1992), *Ontwikkeling van de jeugdcriminaliteit en de justitiële jeugdbescherming: periode 1980-1990* Arnhem: Gouda Quint.

Landgraf JM, Abetz L, Ware JE (1996), *Child Health Questionnaire (CHQ): A Users's Manual.* Boston, MA: The Health Institute, New England Medical Center.

Loeber R, Dishion T, Patterson G (1984), Multiple gating: a multistage assessment procedure for identifying youths at risk for delinquency. *J Res Crime Delinq* 21: 7-32.

Mangunkusumo R, Moorman P, Berg van den-Ruiter de A, Lei van der J, Koning de H, Raat H (2005), Internet-administered adolescent questionnaires compared with a paper version in a randomized study. *J Adolesc Health* 36: 70-76.

Mangunkusumo RT, Duisterhout JS, De Graaff N, Maarsingh EJ, Koning de HJ, Raat H (2006), Internet versus paper mode of health and health behavior questionnaires in elementary schools: asthma and fruit as examples. *J Sch Health* 76: 80-86.

McCabe SE, Boyd CJ, Young A, Crawford S, Pope D (2005), Mode effects for collecting alcohol and tobacco data among 3rd and 4th grade students: a randomized pilot study of web-form versus paper-form surveys. *Addict Behav* 30: 663-71.

Paperny DM, Aono JY, Lehman RM, Hammar SL, Risser J (1990), Computer-assisted detection and intervention in adolescent high-risk health behaviors. *J Pediatr* 116: 456-62.

Raat H, Landgraf JM, Bonsel GJ, Gemke RJBJ, Essink Bot ML (2002), Reliability and validity of the child health questionnaire-child form (CHQ-CF87) in a Dutch adolescent population. *Qual Life Res* 11: 575-581.

Rosenberg M (1965), *Society and the Adolescent Self-image*. Princeton: Princeton University Press.

Rothman KJ, Greenland S (1998), *Modern Epidemiology.* Philadelphia: Lippincott-Raven Publishers.

Serdula MK, Mokdad AH, Pamuk ER, Williamson DF, Byers T (1995), Effects of question order on estimates of the prevalence of attempted weight loss. *Am J Epidemiol* 142: 64-67.

Tourangeau R, Smith TW (1996), Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opin Q* 60: 275-304.

Truman J, Robinson K, Evans AL, Smith D, Cunningham L, Millward R, Minnis H (2003), The Strengths and Difficulties Questionnaire: A pilot study of a new computer version of the self-report scale. *Eur Child Adolesc Psychiatry* 12: 9-14.

Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL (1998), Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science* 280: 867-73.

Van der Linden FJ, Dijkman, T.A., and Roeder, P.J.B. (1983), *Schoolgaande jongeren, hun leefwereld en zelfbeleving.* Nijmegen, The Netherlands: Hoogveldt Instituut.

Van Widenfelt BM, Goedhart AW, Treffers PDA, Goodman R (2003), Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 12: 281-289.

Vereecken CA, Maes L (2006), Comparison of a computer-administered and paper-and-pencil-administered questionnaire on health and lifestyle behaviors. *J Adolesc Health* 38: 426-432.

Wang YC, Lee CM, Lew-Ting CY, Hsiao CK, Chen DR, Chen WJ (2005), Survey of substance use among high school students in Taipei: web-based questionnaire versus paper-and-pencil questionnaire. *J Adolesc Health* 37: 289-95.

Webb PM, Zimet GD, Fortenberry JD, Blythe MJ (1999), Comparability of a computer-assisted versus written method for collecting health behavior information from adolescent patients. *J Adolesc Health* 24: 383-8.

Wright DL, Aquilino WS, Supple AJ (1998), A comparison of computer-assisted paper-and-pencil self-administered questionnaires in a survey on smoking, alcohol, and drug use. *Public Opin Q* 62: 331-353.

Chapter 4

Seasonal variation in self-reported health and health-related behaviour in Dutch adolescents

Petra M. van de Looij-Jansen, Erik Jan de Wilde,
Cathelijne L. Mieloo, Marianne C.H. Donker, Frank C. Verhulst

Accepted as short communication in Public Health 2009

Abstract

Background: **Aim of the present study is to investigate if self-reported (mental) health and health-related behaviour varies by season. The issue of seasonality in youth health surveys has rarely been investigated.**

Methods: **Data were collected within the framework of the Youth Health Monitor Rotterdam (RYM), a longitudinal health surveillance system carried out by the Municipal Public Health Service. Data were collected at schools in the period 2003-2006. A total of 33,171 first and third grade pupils (mainly 12-15 year olds) completed a questionnaire. ANOVA's were used to assess seasonal variation in (mental) health and health-related behaviour.**

Results: **Significant but small seasonal differences were found for psychological well-being, self-esteem, the subscales 'conduct problems', 'hyperactivity', and 'peer problems' of the Strengths and Difficulties Questionnaire (SDQ), the SDQ total problem score, having breakfast, sporting activities, watching television, and alcohol consumption. The strongest effect was found for alcohol consumption, showing the highest scores in May-June. However, all differences remained far below Cohen's criteria for a small effect size.**

Conclusion: **We found no strong evidence for major seasonal variation in self-reported (mental) health and health-related behaviour among adolescents in the Netherlands. A possible implication of our study is that for the design of youth health surveys it is not necessary to spread the data collection over the year.**

Introduction

Youth health surveys are often used to collect data on the prevalence of health and health-related behaviour among youth. Besides varying methods of data collection (e.g. self-administered or interview, paper-and-pencil or web-based) or settings (e.g. school or household), studies also vary in the period or season of data collection. In some surveys data are collected in a specific period of the year, in other surveys data collection is spread throughout the year. Collecting data at different periods of the year raises the issue of seasonal variation in youth health surveys. Knowledge of seasonal variation in health and health-related behaviour is important for the design of (epidemiological) studies and the comparability with data from other youth surveys.

For adults, several studies have investigated seasonal variation in health and health-related behaviour, for example for mental health (Bouwman-Notenboom et al., 2005; de Graaf et al., 2005; Harmatz et al., 2000; Magnusson, 2000), physical activity (Matthews et al., 2001; Pivarnik et al., 2003; Plasqui and Westerterp, 2004), consumption of fruit and vegetables (Cox et al., 2000), alcohol consumption (Cho et al., 2001; Uitenbroek, 1996) and BMI (Shahar et al., 2001; Visscher and Seidell, 2004). Less attention has been given to this subject in youth populations. A few studies exist on mental diagnostic outcomes. Kovalenko et al. assessed seasonal variation in symptoms of internalizing, externalizing, and substance use disorders in American youth (Kovalenko et al., 2000). Significant annual variations were found for overanxious disorder, obsessive-compulsive disorder, separation anxiety disorder, social phobia, and major depressive disorder, with lowest symptom scores in August-October. Weak seasonal variations were found in attention-deficit/hyperactivity disorder, oppositional defiant disorder, and marijuana use. No significant seasonality was found for alcohol, other substance use, agoraphobia or panic disorder. Related to this, the prevalence of Seasonal Affective Disorder (SAD) among youth has been reported in American (Swedo et al., 1995), Swedish (Rastad et al., 2006), Italian (Tonetti et al., 2007) and Finnish samples (Sourander et al., 1999), with percentages ranging from 1.7% (Swedo et al., 1995) to 20.1% (Rastad et al., 2006).

Furthermore, a few studies investigated seasonality in health-related behaviours. In a study of Colwell et al. seasonal variation in the initiation of smoking among American adolescents (enrolled in a Tobacco Awareness & Cessation Program) was found (Colwell et al., 2006). Youth were significantly more likely to begin smoking during the summer. Conversely, the number of youth beginning to smoke during the fall semester at school as well as during the winter season was significantly lower than expected. In Canadian youth (n=200, age 8-16 yrs) Crocker et al. found that activity patterns were higher in spring compared with those in late fall and winter (Crocker et al., 1997). Seasonal variation in the dietary intake of fruits and grains was found in a study of Givens et al. among a small sample of American youth (n=23, 3-11yrs) (Givens et al., 2007). The highest consumption of fruits was found in summer, whereas the highest consumption of grains was found in winter. For vegetables and juice no significant differences were found.

The explanation for seasonal variation lies in differences in climate (e.g. length of day, hours of sunshine), holidays or social contacts. For mood, annual variation is thought to be explained by annual variation in brain serotonin, modulated by exposure to bright light (Arora et al., 1984; Rao et al., 1990). Therefore, due to differences in climate or culture, the results of the aforementioned studies cannot be generalized to youth in the Netherlands. To our knowledge, no large-scale epidemiological surveys have been conducted to investigate seasonality in health or health-related behaviour among adolescents in the Netherlands. Therefore, the aim of the present study is to assess this variation.

If this seasonal pattern indeed exerts influence on the health and health-related behaviour, this would imply that the research methods should take this factor into account. For example, the fieldwork of large population studies may have to be spread throughout the year and weighting for the proportion of respondents per season should be applied in order to render the data representative of a whole year.

Methods

Study design and procedure

Data were obtained from the Rotterdam Youth Monitor (RYM), a longitudinal youth health surveillance system carried out by the Municipal Public Health Service. Main aim of the RYM is to monitor the general health, well-being, behaviour, and related factors of youth aged 0-19 years living in Rotterdam (the Netherlands) and surroundings in order to supply information for youth policy at school, neighbourhood and municipality level. The RYM is incorporated into the regular health examinations of the preventive youth healthcare system, which means that RYM data are also used to detect potential individual health risks and problems in order to take necessary preventive measures (including referrals for treatment).

At secondary schools health examinations are offered to the ages 12-13 year-olds (first graders), and 14-15 year-olds (third graders). Data are collected throughout the school year, except for the months July and August (because of summer holidays). In practice, most questionnaires are administrated in the period November-December, because all the pupils who complete a questionnaire have to be invited for a routine health examination during the school year. About 85% of all secondary schools in Rotterdam participate in the RYM.

The present study used data on first (mainly 12-13 year olds) and third-grade pupils (mainly 14-15 year olds) from secondary schools. These pupils completed a RYM questionnaire between September 2003 and July 2004 (n=12,795), September 2004 and July 2005 (n=10,141), or September 2005 and July 2006 (n=10,235). Specially trained school nurses of the Municipal Public Health Service carried out the administration of the questionnaire during regular class hours. Details about the administration procedure (date of administration, questions asked, reasons for absence, time needed to fill in the questionnaire) were noted on a special form.

All data were gathered within and as part of the government approved routine health examinations of preventive youth healthcare. Separate informed consent was therefore not requested. Only anonymous data were used. The questionnaires were completed on a voluntary basis. Parents received written information on these questionnaires and were free to object to their child's participation.

Subjects

A total of 33,171 pupils completed the questionnaire (response rate 90%). The reason for non-response was mainly illness. As the actual date of data collection was missing for 42 cases, the data of 33,129 pupils were included in the analysis.

Measures

Season

The administration dates of the questionnaires were categorised into five periods: September-October, November-December, January-February, March-April, and May-June.

The Netherlands is situated on the North Sea and has a mild maritime climate. Summers are not extremely hot, nor are winters extremely cold. Mean hours of sunlight in the period of data collection ranged from 63.3 hours in November-December to 226.3 hours in May-June (data provided by the Royal Netherlands Meteorological Institute, 2007).

(Mental) health

Perceived health was assessed by means of one question in which the respondents were asked to rate their health (CBS, 2001) Respondents reported their answers on a 5-point scale ranging from very good (1) to bad (5).

Psychological well-being was measured with nine items about feelings and moods, derived from the Child Health Questionnaire (CHQ-CF87): four items about positive feelings (feeling happy, feeling cheerful, enjoying the things you do, having fun) and five items about negative feelings (feeling lonely, feeling sad, feeling afraid or scared, worrying about things, feeling unhappy) (Landgraf et al., 1996; Raat et al., 2002) Students were asked how much of the time these feelings were present during the past four weeks. Each item was rated on a 5-point scale ranging from 'very often' to 'never'. Scores range from 0 to 100, with a high score indicating positive psychological well-being. Cronbach's α in this sample was 0.86.

Self-esteem was measured by the Dutch version of the Rosenberg Self-esteem Scale (Rosenberg, 1965; van der Linden, 1983) This scale consists of ten statements (e.g. "I feel that I have a number of good qualities"). Each item was answered on a 4-point scale (strongly disagree, disagree, agree and strongly agree). Scores range from 0 to 10, with a high score indicating higher self-esteem. Cronbach's α in this sample was 0.83.

Mental health was measured by the Dutch self-report version of the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 2001; van Widenfelt et al., 2003). The SDQ consists of 25

items that can be allocated to five subscales of five items each: emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour. All items refer to the past six months. Each item has to be scored on a 3-point scale with 0='not true', 1='somewhat true', and 2='certainly true'. Subscale scores were computed by summing scores on relevant items (after recoding reversed items; range 0-10). A total difficulties score can be calculated by summing the scores on the emotional symptoms, conduct problems, hyperactivity-inattention, and peer problems subscales (k=20; range 0-40). Cronbach's α in this sample was 0.71.

Health-related behaviour

Fruit and vegetable consumption was measured by the questions "How many days per week do you eat fruit (or vegetables)?", with answers ranging from 0 days to 7 days. Having breakfast was measured by the question "How many days per week do you have breakfast?" (answers ranging from 0 days to 7 days).

Sporting activities was assessed by the question "How many days per week do you play a sport?" and watching television by the question "How many hours per day do you watch television?" (answers ranging from 0 to 5).

For alcohol consumption we used the question concerning frequency of consumption in the past four weeks; possible answers ranged from '0 times' (1) to 'more than 10 times' (6). Cannabis use was measured by the question "How often have you used cannabis in the past four weeks?", with possible answers ranging from '0 times' (1) to 'more than 10 times' (6).

Demographic variables

Demographic questions included gender, age, level of education, country of birth and parents' country of birth. Level of education was dichotomised into the categories 'low education' (including basic pre-vocational education and theoretical pre-vocational education) and 'high education' (including general secondary education and pre-university education). Age was dichotomised into the categories 'below 14 years and '14 years or older'. Ethnic background was classified as Dutch or non-Dutch. According to definitions of Statistics Netherlands, adolescents with at least one parent born outside the Netherlands were classified as non-Dutch.

Data analysis

One-way analysis of variance was performed to study differences in health and health-related behaviour across the seasons. Season (five periods) was entered into the analysis as independent variable. All analyses were adjusted for gender, age, level of education, ethnic background and year of data collection. Games-Howell post-hoc testing was carried out in the case of significant differences across season. Gender-by-season interactions were entered into the analysis to determine whether seasonality varied by gender. According to Cohen's criteria, effect sizes of 0.10, 0.25 and 0.40 were considered to represent small, medium and large differences, respectively (Cohen, 1988).

Results

Table 4.1 shows the distribution of socio-demographic characteristics across the two-month periods. For all socio-demographic characteristics significant differences were found across the period of data collection.

Table 4.1 Socio-demographic characteristics of the study population (n=33,129), by period of data collection

	Sept-Okt n=6438	Nov-Dec n=13,777	Jan-Febr n=6350	March-April n=4695	May-June n=1869	χ^2 (df)	p
Gender							
boys	52.4%	52.3%	48.3%	50.7%	47.1%	44.00 (4)	‹ 0.001
girls	47.6%	47.7%	51.7%	49.3%	52.9%		
Age							
‹ 14 years	42.6%	56.8%	43.0%	42.7%	26.2%	970.77 (4)	‹ 0.001
≥ 14 years	57.4%	43.2%	57.0%	57.3%	73.8%		
Level of education							
low	63.7%	60.0%	59.0%	52.8%	62.6%	145.15 (4)	‹ 0.001
high	36.3%	40.0%	41.0%	47.2%	37.4%		
Ethnic background							
Dutch	41.9%	51.5%	46.5%	52.3%	59.7%	283.27 (4)	‹ 0.001
non-Dutch	58.1%	48.5%	53.5%	47.7%	40.3%		

Table 4.2 presents the mean scores (sd) for the different health indicators and the results of the analysis of variance.

Table 4.2 Means (SD) of health indicators; association with season (ANOVA: F-values, effect sizes, post hoc analysis)

	Period	Mean	SD
(Mental) health			
perceived health (1-5)	actual	1.94	0.72
psychological well-being (0-100)#	past 4 weeks	79.03	14.86
self-esteem (0-10)#	actual	7.98	2.14
Strengths and Difficulties Questionnaire (SDQ)	past 6 months		
- SDQ emotional symptoms (0-10)		2.59	2.10
- SDQ conduct problems (0-10)		2.07	1.56
- SDQ hyperactivity (0-10)		3.73	2.32
- SDQ peer problems (0-10)		2.07	1.71
- SDQ prosocial behaviour (0-10)		7.35	1.83
- SDQ total difficulties score (0-40)		10.45	5.01
Health-related behaviour			
fruit consumption, days per week (0-7)	actual	4.66	2.12
vegetable consumption, days per week (0-7)	actual	5.19	1.77
having breakfast, days per week (0-7)	actual	5.38	2.32
sporting activities (0-7)#	actual	3.07	2.13
watching television (0-5), hours per day	actual	2.66	1.06
use of alcohol (1-6)	past 4 weeks	1.47	0.89
use of marijuana (1-6)	past 4 weeks	1.09	0.51

[1] *All analysis are adjusted for gender, age, educational level, ethnic background, and year of data collection*

** p ‹ 0.05, ** p ‹ 0.01, *** p ‹ 0.001; # data collected in 2004-2005 and 2005-2006*

F-value[1]	Effect size (f)	Post Hoc analysis
2.05	-	
3.38 **	0.03	March/April ‹ Sept/Oct, Nov/Dec, Jan/Feb
3.07*	0.03	pairwise differences not statistically significant
1.85	-	
6.11***	0.03	Nov/Dec ‹ Jan/Feb, May/June
8.12***	0.03	May/June › other months
5.27***	0.03	Sept/Oct › Nov/Dec, Jan/Feb, March/April
2.03	-	
8.09***	0.03	May/June › other months
1.04	-	-
2.07	-	-
2.38*	0	
5.02**	0.03	Jan/Feb ‹ Sept/Oct, Nov/Dec, March/April, May/June
8.89***	0.03	May/June › Sept/Oct, Nov/Dec, March/April
13.64***	0.04	May/June › other months
1.48	-	

(Mental) health

For the variables psychological well-being, self-esteem, the SDQ subscales 'conduct problems', 'hyperactivity', and 'peer problems', and the SDQ total difficulties score, significant differences were found per period of data collection. However, effect sizes did not reach the Cohen's criteria for a small effect size. Figure 4.1 shows the mean scores of the SDQ total difficulties score per season, with highest scores in May-June. No significant seasonal differences were found for the SDQ subscales 'emotional symptoms' and 'prosocial behaviour', and perceived health.

Health-related behaviour

Significant seasonal differences were found for having breakfast, sporting activities, watching television, and alcohol consumption. All effect sizes remained below the Cohen's criteria for a small effect size. The strongest effect was found for alcohol consumption, showing the highest scores in May-June (Figure 4.2). For fruit and vegetable consumption, and use of marijuana no significant seasonal variation was found.

Gender-by-season interaction *(data not shown)*

Significant gender-by-season interactions were found for SDQ emotional symptoms ($F=3.01$, $p<0.05$, effect size 0), SDQ peer problems ($F=4.06$, $p<0.01$, effect size 0) and watching television ($f=2.59$, $p<0.05$, effect size 0).

Figure 4.1 Mean Total Difficulties Score SDQ (95% CI), by months of data collection

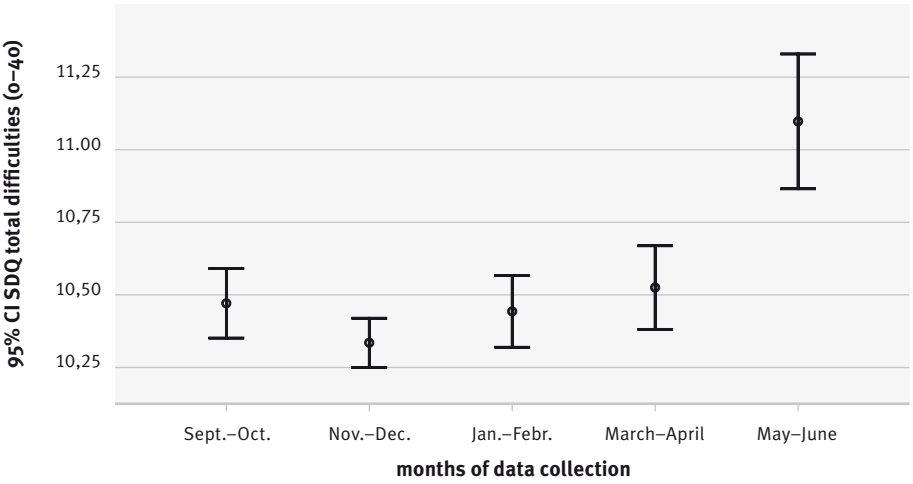
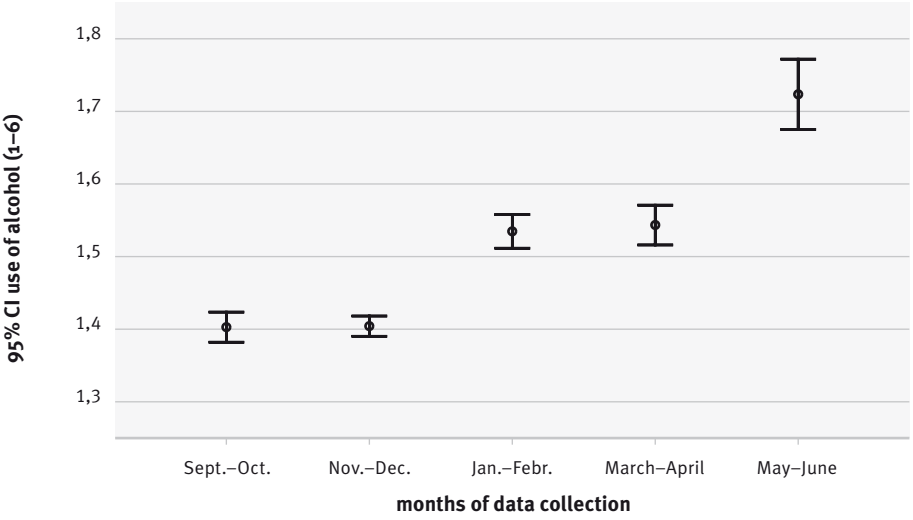


Figure 4.2 Mean use of alcohol (95% CI), by months of data collection



Discussion

In this large school-based study we examined seasonal variation in self-reported (mental) health and health-related behaviour among mainly 12-15 year old adolescents. There were significant seasonal differences in psychological well-being, self-esteem, the SDQ subscales 'conduct problems', 'hyperactivity', and 'peer problems', the SDQ total problem score, having breakfast, sporting activities, watching television, and alcohol consumption. Gender-by-season differences were hardly found. Effect sizes of these differences appeared to be negligible, as all effect sizes remained far below the criteria of Cohen for a small effect size. The strongest effect, but still small, was found for alcohol consumption, showing the highest scores in May-June. Furthermore, for comparison, the difference between boys and girls in emotional symptoms was 0.22 in terms of effect size (a medium effect size) in this sample. This difference is five times larger than the largest seasonal difference in the present study. A critical remark towards previous studies may be that it is remarkable that none of these calculated or mentioned effect sizes.

However, comparisons with previous studies remain limited, as few studies examined seasonal variation in health and health-related behaviour among youth at all. With regard to health-related behaviours, on the subject of sporting activities the present results are in line with the findings of Crocker et al. (1997), showing the highest scores in spring. For alcohol, we found the highest consumption in May-June. This is not in line with the results of Kovalenko et al. (2000) who found no seasonal differences in alcohol consumption, but our results do follow the pattern of adult studies (Cho et al., 2001; Uitenbroek, 1996). Different activities and vacations during May-June may be a possible explanation for this finding. For both fruit and vegetable consumption we found no significant seasonal differences. Givens et al. (2007) found seasonal variation in fruit consumption (highest consumption in summer), but not in vegetable consumption. A possible explanation is that these differences are likely to be related to the year-round availability of fruit in the Netherlands.

Despite the few studies that found evidence for SAD in adolescents (Rastad et al., 2006; Sourander et al., 1999; Swedo et al., 1995; Tonetti et al., 2007) and the results of Kovalenko et al. (2000) who found annual variation for internalizing disorders, this study found no major seasonal variation for mental health characteristics in Dutch youth. This might be explained by the fact that SAD was not measured directly. However, a disadvantage of specific instruments measuring SAD (e.g. the Seasonal Pattern Assessment Questionnaire) is that these instruments rely on the participants' own subjective evaluation of their seasonal difficulties, and also on their retrospective recall of the temporal pattern. Another possible explanation for the limited seasonal differences we found in our study may be that differences in climate in the Netherlands are probably much smaller compared with other countries. A third possibility to explain the lack of effect could be that other factors, such as an increase of school-related stress at the end of the school term, have effect in the opposite direction. As this is the first time that seasonal variation has been studied in a Dutch adolescent sample, no comparable national data for seasonal variation in youth health are available.

Furthermore, the administration dates of the questionnaire were divided into five periods of two months each. As one could argue this is an arbitrary choice, the data were also analysed with a division into three periods (September/October/November, December/January/February, and March/April/May/June). However, no substantial differences were found in comparison with the results reported when five periods were chosen.

One of the strengths of our study, although not designed to study seasonal differences, is that it was possible to use data over a period of three years. In this way, our results are probably less influenced by coincidental factors in a specific year, such as special events or annual differences in temperature.

A possible limitation is this was a school-based study. As a consequence, no data were collected in July and August because of the school holidays. However, data collection in July and August might always be problematic, regardless the method of data collection, due to summer holidays. For instance, a mailed questionnaire might result in a low and probably selective response rate. A remark is that for those health indicators reported for the 'past 6 months' the school holiday period (July and August) was included.

In the Netherlands, governments are obliged to monitor the health of young people. Therefore, most Municipal Public Health Services periodically carry out a youth health survey, usually a self-report questionnaire that is administrated in classrooms. Until now, seasonal variation in health indicators among youth was assumed, although strong evidence was lacking. As the results of our study show, no major seasonal variation in self-reported health was found. A possible implication of our study is that for the design of youth health surveys it is not necessary to spread the data collection throughout a one-year period.

In conclusion, no strong evidence was found for major seasonal differences in health and health-related behaviour among adolescents in a country with a mild maritime climate. More research in youth is needed to investigate whether this is true for countries with larger climatic differences.

References

Arora RC, Kregel L, Meltzer HY (1984), Seasonal variation of serotonin uptake in normal controls and depressed patients. *Biol Psychiatry* 19: 795-804.

Bouwman-Notenboom AJ, Toet J, van Ameijden EJC (2005), Seizoensinvloeden in gezondheidsenquêtes. *Tijdschr Gezondh* 83: 335-340.

CBS (2001), *Jeugd, cijfers en feiten* Voorburg/Heerlen: CBS.

Cho YI, Johnson TP, Fendrich M (2001), Monthly variations in self-reports of alcohol consumption. *J Stud Alcohol* 62: 268-72.

Cohen J (1988), *In Statistical power analysis for the behaviour sciences (Rev. edn.)* New York: Academic.

Colwell B, Ramirez N, Koehly L, Stevens S, Smith DW, Creekmur S (2006), Seasonal variations in the initiation of smoking among adolescents. *Nicotine Tob Res* 8: 239-43.

Cox BD, Whichelow MJ, Prevost AT (2000), Seasonal consumption of salad vegetables and fresh fruit in relation to the development of cardiovascular disease and cancer. *Public Health Nutr* 3: 19-29.

Crocker PR, Bailey DA, Faulkner RA, Kowalski KC, McGrath R (1997), Measuring general levels of physical activity: preliminary evidence for the Physical Activity Questionnaire for Older Children. *Med Sci Sports Exerc* 29: 1344-9.

De Graaf R, van Dorsselaer S, ten Have M, Schoemaker C, Vollebergh WA (2005), Seasonal variations in mental disorders in the general population of a country with a maritime climate: findings from the Netherlands mental health survey and incidence study. *Am J Epidemiol* 162: 654-61.

Garnefski N, Diekstra RFW (1993), *Scholierenonderzoek 1992: landelijke uitkomsten rond een onderzoek naar gedrag en gezondheid van ca. 11.000 scholieren van het voortgezet onderwijs.* Leiden: Rijksuniversiteit Leiden.

Givens ML, Lu C, Bartell SM, Pearson MA (2007), Estimating dietary consumption patterns among children: a comparison between cross-sectional and longitudinal study designs. *Environ Res* 103: 325-30.

Goodman R (2001), Psychometric properties of the Strengths and Difficulties Questionnaire. *J Am Acad Child Adolesc Psychiatry.* 40: 1337-1345.

Harmatz MG, Well AD, Overtree CE, Kawamura KY, Rosal M, Ockene IS (2000), Seasonal variation of depression and other moods: a longitudinal approach. *J Biol Rhythms* 15: 344-50.

Kovalenko PA, Hoven CW, Wicks J, Moore RE, Mandell DJ, Liu H (2000), Seasonal variations in internalizing, externalizing, and substance use disorders in youth. *Psychiatry Res* 94: 103-19.

Landgraf JM, Abetz L, Ware JE (1996), *Child Health Questionnaire (CHQ): A Users's Manual.* Boston, MA: The Health Institute, New England Medical Center.

Magnusson A (2000), An overview of epidemiological studies on seasonal affective disorder. *Acta Psychiatr Scand* 101: 176-84.

Matthews CE, Freedson PS, Hebert JR, Stanek EJ, 3rd, Merriam PA, Rosal MC, Ebbeling CB, Ockene IS (2001), Seasonal variation in household, occupational, and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study. *Am J Epidemiol* 153: 172-83.

Pivarnik JM, Reeves MJ, Rafferty AP (2003), Seasonal variation in adult leisure-time physical activity. *Med Sci Sports Exerc* 35: 1004-8.

Plasqui G, Westerterp KR (2004), Seasonal variation in total energy expenditure and physical activity in Dutch young adults. *Obes Res* 12: 688-94.

Raat H, Landgraf JM, Bonsel GJ, Gemke RJBJ, Essink Bot ML (2002), Reliability and validity of the child health questionnaire-child form (CHQ-CF87) in a Dutch adolescent population. *Qual Life Res* 11: 575-581.

Rao ML, Muller-Oerlinghausen B, Mackert A, Stieglitz RD, Strebel B, Volz HP (1990), The influence of phototherapy on serotonin and melatonin in non-seasonal depression. *Pharmacopsychiatry* 23: 155-8.

Rastad C, Ulfberg J, Sjoden PO (2006), High prevalence of self-reported depressive mood during the winter season among Swedish senior high school students. *J Am Acad Child Adolesc Psychiatry* 45: 231-8.

Rosenberg M (1965), *Society and the Adolescent Self-image*. Princeton: Princeton University Press.

Shahar DR, Yerushalmi N, Lubin F, Froom P, Shahar A, Kristal-Boneh E (2001), Seasonal variations in dietary intake affect the consistency of dietary assessment. *Eur J Epidemiol* 17: 129-33.

Sourander A, Koskelainen M, Helenius H (1999), Mood, latitude, and seasonality among adolescents. *J Am Acad Child Adolesc Psychiatry* 38: 1271-6.

Swedo SE, Pleeter JD, Richter DM, Hoffman CL, Allen AJ, Hamburger SD, Turner EH, Yamada EM, Rosenthal NE (1995), Rates of seasonal affective disorder in children and adolescents. *Am J Psychiatry* 152: 1016-9.

Tonetti L, Barbato G, Fabbri M, Adan A, Natale V (2007), Mood seasonality: a cross-sectional study of subjects aged between 10 and 25 years. *J Affect Disord* 97: 155-60.

Uitenbroek DG (1996), Seasonal variation in alcohol use. *J Stud Alcohol* 57: 47-52.

Van der Linden FJ, Dijkman, T.A., and Roeder, P.J.B. (1983), *Schoolgaande jongeren, hun leefwereld en zelfbeleving*. Nijmegen: Hoogveldt Instituut.

Van Widenfelt BM, Goedhart AW, Treffers PDA, Goodman R (2003), Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 12: 281-289.

Visscher TL, Seidell JC (2004), Time trends (1993-1997) and seasonal variation in body mass index and waist circumference in the Netherlands. *Int J Obes* 28: 1309-16.

Chapter 5

Differences in measured and self-reported height and weight in Dutch adolescents

Wilma Jansen, Petra M. van de Looij-Jansen, Isabel Ferreira, Erik Jan de Wilde, Johannes Brug

Annals of Nutrition and Metabolism 2006, 50: 339-346

Abstract

***Background*: Monitoring overweight prevalence and its trends in Dutch youth is frequently based on self-reported data. The validity of self-reported data especially in young adolescents is not sufficiently known. The purpose of this study is to study the validity of self-reported height and weight in 12-13 year olds, to identify socio-demographic correlates and to explore whether correction factors can be developed to estimate the prevalence of overweight in youth.**

***Methods*: 5.525 12-13 year old pupils the Rotterdam area filled in a confidential questionnaire on health topics, including their height and weight. In a sub sample of 499 pupils both self-reported and measured height and weight were available.**

***Results*: Self-reported data led to a considerable underestimation of Body Mass Index and consequently the prevalence of overweight. Underestimation was higher in pupils who regarded themselves as more fat, were from non Dutch origin and in lower education levels.**

***Conclusions*: Self-reported height and weight appeared to be inappropriate to estimate the overweight prevalence in 12-13 year olds, unless the figures were adjusted. Using adjusted self-reported BMI on an individual level is questionable. Actual measurements of height and weight are necessary to draw up valid correction formulas in new samples.**

Introduction

In the Netherlands, as in most countries in the developed and developing world, the prevalence of overweight and obesity among children and adolescents is increasing (Deckelbaum and Williams, 2001; Lobstein and Frelut, 2003). A national survey showed that in the Netherlands this prevalence increased by about 50% from 10% in 1980 to 15% in 1997 (Hirasing et al., 2001). Monitoring the prevalence of overweight and obesity, and its trends in the general population has therefore been adopted by several youth health care authorities in the Netherlands. Goals are to identify associated underlying risk factors and sub-groups at higher risk, to be targeted in preventive intervention measures and to (re)direct health policy decisions. However, many of these large-scale monitoring systems rely on self-reported data on weight and height, based on which the prevalence of overweight and obesity in the population under study is estimated. The question that therefore arises is to what extent self-reported methods are valid tools for estimating the prevalence of overweight and obesity and its trends in populations of children and/or adolescents. Another question is to what extent these self-reported data are valid tools to identify individuals with overweight, so they can be targeted for individual interventions. Measuring height and weight of large groups of individuals is costly and if self-reports are valid, one could use these to inform interventions tailored to weight status (Brug et al., 2003).

A review of the literature on the validity of self-reported height and weight in children and adolescents (Table 5.1) suggests that, in most cases, height is over-reported, whereas weight is under-reported, thereby leading to an underestimation of the prevalence of overweight and obesity (Brener et al., 2003; Brooks-Gunn et al., 1987; Crawley and Portides, 1995; Davis and Gergen, 1994; Farré Rovira et al., 2002; Fortenberry, 1992; Giacchi et al., 1998; Goodman et al., 2000; Hauck et al., 1995; Himes and Faricy, 2001; Himes et al., 2005; Shannon et al., 1991; Strauss, 1999; Tienboon et al., 1992). The differences between self-reports and objective assessments may be associated with several factors such as gender (Brener et al., 2003; Crawley and Portides, 1995; Goodman et al., 2000), age (Brener et al., 2003; Crawley and Portides, 1995; Davis and Gergen, 1994; Himes et al., 2005), ethnicity (Brener et al., 2003; Himes et al., 2005), socio-economic status (Himes et al., 2005) and overweight/obesity status (Crawley and Portides, 1995; Davis and Gergen, 1994; Fortenberry, 1992; Himes et al., 2005; Strauss, 1999), although no consistent pattern appears from earlier studies. Upon the usefulness of self-reported data different conclusions are drawn.

In many of the reviewed studies self-report was followed almost directly by an actual measurement, with subjects being aware of this procedure (Brener et al., 2003; Farré Rovira et al., 2002; Goodman et al., 2000; Himes and Faricy, 2001; Strauss, 1999). This might well have influenced the accuracy of the self-reported data and the conclusions drawn from these studies might not be valid for monitor systems which rely solely on self-report data.

Table 5.1 Review of literature on validity of self-reported height and weight by adolescents.

Reference	Population	Country	Age	Sample type	Later measurement known to subjects?
Brener et al., 2003	4.619 high school students	USA	14-19 y	convenience	yes
Brooks-Gunn et al., 1987	151 girls	USA	11-13 y	not representative, high SES	?
Crawley and Portides, 1995	1240 adolescents	UK	16-17 y	sub sample of national cohort study	?
Davis and Gergen, 1994	829 Mexican-American adolescents	USA	12-19 y	representative of 5 states	?
Farré Rovira et al., 2002	568 adolescents	Spain	14-20 y	representative for region	yes
Fortenberry, 1992	725 adolescent patients	USA	14-20 y	clinical	?
Giacchi et al., 1998	133 secondary school students	Italy	15-17y	random sample of schools in Siena	?
Goodman et al., 2000	15.000 teenagers	USA	12-19 y	national longitudinal	yes
Hauck et al., 1995	806 Indian Americans	USA	12-19 y	convenience sample	?
Himes and Faricy, 2001	14 million adolescents	USA	12-17 y	national representative (Nhanes 3)	yes
Himes et al., 2005	3797 students .	USA	12-18 y	project participants	?

% overweight	Difference between self report and measured Height	Weight	BMI	Correlates of these differences	Conclusion on validity
28% (› P85)	6.9 cm	-1.6 kg	-2.6	girls underreport more, higher grades under-report more	positive
?	+1.4 cm	-1.2 kg	-	-	inconclusive
8% for males, 12% for females (›25 kg/m²)	+1.1 cm for males, +0.8 for females	+0.4 for males, -0.9 for females	-0.1 for males, -0.5 for females	-	-
?	-0.1% for males, 0.4% for females	+0.8% for males, -1.8% for females	1.3% for males, -2,4% for females	high BMI groups underreport BMI more than low BMI groups	positive
23% (BMI›25)	1.5 cm for males, 2.5 cm for females	-2 kg for males, -1 kg for females	-		inconclusive
10% (›P90)	0.6 cm for males, 0.5 cm for females	-1.2 kg for males, -1.5 kg for females	-	high BMI groups underreport BMI more than low BMI groups	negative
17% for males, 12% for females (BMI›25)	0.8 cm for males, 2.3 cm for females	–1.8 kg for males, -1.9 kg for females	-0.8 for males, -1.3 for females		-
10% (›P95)	-	-0,2 kg for males, -1,0 kg for females	+0,3 for males, -2,7 for females	girls underre-port BMI more than boys.	positive
34-44% (›P85)	1.5 cm for males, 0.5 cm for females	-1.5 kg for males, -2.1 kg for females	-	-	negative
?	-1.0 cm for males, -1.1 cm for females	-0.5 kg for males, -1.2 kg for females	+0.2 for males, -0.1 for females	-	negative for age ‹ 14 y
33% (›P85)	1.2 cm for males, 2.4 cm for females	-1.6 for males, - 3.5 for females	-	-	positive

Continue on next page

Table 5.1 Review of literature on validity of self-reported height and weight by adolescents. *(Continue)*

Reference	Population	Country	Age	Sample type	Later measurement known to subjects?
Shannon et al., 1991	806 Caucasian sixth graders	USA	11-12 y	-	?
Strauss, 1999	1932 adolescents	USA	12-16 y	national representative (Nhanes 3)	yes
Tienboon et al., 1992	204 adolescents	Australia	14-15 y	sub sample of prospective study	no

Most studies on validity of self-reported height and weight have been conducted in the USA. Only a limited number of studies were performed in Europe (UK, Spain, Italy). In the Netherlands, the validity of self-reported height and weight measures among adolescents has never been evaluated, despite its use by several health care surveillance systems. Due to different population characteristics (for example ethnicity and overweight prevalence) the validity of self-report might not be the same in every population.

The number of studies that included self-reports of 12-13 year olds is also limited, while this age group is an official target age group in Dutch preventive youth health care.

The purpose of the present study is therefore three-fold: first, to study the validity of self-reported height and weight among a sample of Dutch 12-13 year olds; second, to identify socio-demographic correlates of discrepancies between the self-reported and the measured body dimensions; and third, to explore whether correction factors can be developed to adjust overweight and obesity estimates based on self-reports in youth.

		Difference between self report and measured			Correlates of these	Conclusion
	% overweight	Height	Weight	BMI	differences	on validity
	?	+0.5 cm for males, -0.5 cm for females	-2.1 kg for males, -4.0 kg for females	-	-	negative
	27% (>P85)	-1.0 cm	0.1 kg for males, -1.0. kg for females	-	-	positive
	?	0.8 cm for males, 1.5 cm for females	-0.9 kg for males, -0.6 kg for females	-0.3 for males, -0.6 for females	-	positive

Methods

Study design and subjects

The study sample was drawn from the ongoing Rotterdam Youth Monitor (RYM) Study. This health surveillance system monitors the general health, well-being, behaviour and related factors of youth aged 0 to 19 years living in Rotterdam and surroundings in order to supply information for youth policy at the school, neighbourhood and municipality level. Because the RYM is incorporated in the regular check ups of the preventive youth health care system, the RYM data are also used to detect potential individual health risks and problems in order to take necessary preventive measures (including referrals for treatment).

In the 2003/2004 survey, 48 secondary schools (response rate 86%) in the Rotterdam area were visited by a school nurse of the Municipal Public Health Service who administered the RYM questionnaire during class hours to all first grade pupils. The questionnaire included questions on perceived health status, well-being and body image, and also on body height and weight. Pupils were told their answers would be treated confidentially, they would consult with the school nurse later on in the school year and the school nurse would not have access to their answers. Pupils were not told that they could be measured and weighted by the school nurse. A total of 5525 (93% response rate) answered the questionnaire, of whom 3939 provided self-reports on both height (expressed in cm) and weight (in kg); however among 66 pupils these answers were not plausible (height‹120 cm or ›220 and/or weight‹30kg or ›200kg), and were therefore excluded from further analyses. Multiple logistic regression analysis indicated that pupils with missing self-reported data on height and weight were on average older, from lower levels of education, perceived themselves more often as being overweight, and were more often from Moroccan origin.

About 3.2 (± 1.1) months later pupils were consulted by the school nurse. Three of the 29 school nurses were asked to measure height and weight of all the pupils who attended their consultations. The other nurses decided for themselves whether to measure pupils or not. For 499 pupils both self-reported and measured height and weight were obtained. These pupils constitute the sub-sample analysed in the current study; a sub-sample in which children from non-Dutch origin (71% versus 54%), from lower education levels (53% versus 28%), and with higher self-reported BMI's (19.5 ± 3.7 versus 18.5 ± 2.9) were somewhat over-represented as compared to the Rotterdam population of 12-13-year-olds at large.

Measures

Self-reported and measured height were reported to the nearest cm; self-reported and measured weight to the nearest kg. SECA roll up measuring bands were used for measuring height and calibrated SECA flat weighing scales for weight. Weight was measured without heavy clothing and shoes and was therefore corrected for light clothing by subtracting 1 kg.

Body Mass Index (BMI) was calculated based on self-reported as well as measured length and weight. Weight status (not overweight, overweight) was determined according to the international guidelines presented by Cole et al. (Cole et al., 2000).

Country of origin was determined by country of birth of mother and father. If both parents had been born in the Netherlands, the child's country of origin was Dutch; if one or both parents were born in another country, the country of origin was defined according to that country; if both parents had been born in different foreign countries, the country of origin of the mother was deemed most important. Children were subsequently allocated to the following categories: Dutch, Surinam, Dutch Antillean or Aruban; Moroccan; Turkish; and other.

Body image was measured with one questionnaire item on a five-point scale ('far too thin', 'too thin', 'not thin / not fat', 'too fat', 'far too fat').

Pupils were classified into three different levels of education: 'basic pre-vocational training', 'theoretical pre-vocational training', 'general secondary education' (including junior and senior general secondary education and pre-university training) and three age groups 11-12, 13 and 14 years of age.

Data analyses / statistical methods

Differences in self-reported and measured height, weight and BMI were first explored with paired t-tests. Subsequently, hierarchal multiple linear regression analyses were conducted with measured BMI (BMIm) as dependent variable, self-reported BMI (BMIs) as well as time interval (in month) between the self-report and the actual measurement entered as first block of independents; socio-demographic variables as second block and body image as final block. The correlation between BMIs and body image was not higher than 0.6. Collinearity statistics were checked and the condition index did not exceed acceptable values. Country of origin was dummy-coded with Dutch as reference category. After assessing the main effects, potential modification effects of all socio-demographic independent variables were examined by adding interaction terms with self-reported BMI to the regression model.

Correction equations (including BMIs and socio-demographic variables and optionally body image) were derived from the regression model in order to estimate BMI from BMIs in the total population. Sensitivity and specificity of overweight status based on BMIs and corrected BMIs were calculated and expressed as percentages.

Results

Socio-demographic and selected characteristics of the study population are given in table 5.2.

Table 5.2 Selected and socio-demographic characteristics of sub sample (with self-reported and measured data) and total sample (with self-reported data).

	Sub sample n=499 %	Total sample n= 5525 %
Gender		
Male	53	51
Female	47	49
Age		
11	‹ 1	1
12	50	47
13	44	45
14	5	7
Level of education		
Basic pre-vocational training	51	30
Theoretical pre-vocational training	16	13
General secondary education	33	57
Country of origin		
The Netherlands	29	46
Surinam	15	12
Dutch Antilles or Aruba	11	4
Morocco	13	10
Turkey	16	12
Other	16	16
Body image		
Far too thin	2	2
Too thin	9	11
Not thin / not fat	56	58
Too fat	27	24
Far too fat	6	5
Percentage overweight		
Based on self-report	18	13*
Based on measurement	33	-

**: percentage based on n=3939 due to missing values*

Self-reported height was 1.5 cm lower than measured height, self-reported weight 6.1 kg lower than measured weight, resulting in an underestimation of BMI of 1.5 kg/m². T-tests showed that all differences were significant at the $p < 0.001$ level (Table 5.3). Correlations between self-reports and measurements ranged from 0.75 for BMI and 0.80 for weight to 0.85 for height. The prevalence of overweight in the study population based on self-reported data would be estimated as 18%, while the prevalence of overweight based on measured data appeared to be 33%. Mean differences in self-reported and measured height, weight and BMI by socio-demographic variables, body image, weight status and time interval between self-report and actual measurement are given in table 5.4.

Table 5.3 **Self-reported and measured height, weight and BMI (mean and 95% confidence interval), and Pearson correlations between both assessments.**

	Self-report	Measured	Difference	Pearson's R
Height	159.6 cm	161.2 cm	-1.5 cm	0.80
	(158.8-160.3)	(160.5-161.8)	(-2.0- -1.1)	
Weight	49.7 kg	55.8 kg	-6.1 kg	0.85
	(48.8-50.7)	(54.7-57.0)	(-6.7- -5.5)	
BMI	19.5 kg/m²	21.0 kg/m²	-1.5 kg/m²	0.75
	(19.2-19.8)	(20.6-21.3)	(-1.8- -1.3)	

All differences between self-reports and measured data were significant at the p<0.001 level.

Table 5.4 Mean difference (SD) between self-reported and measured height, weight and BMI by socio-demographic characteristics, body image, measured weight status and time interval between self-report and measurement.

Mean difference (SD) in:	Height (cm)	Weight (kg)	BMI (kg/m²)
Gender			
Male	-2.0 (4.9)	-6.5 (6.9)	-1.6 (2.8)
Female	-1.1 (5.6)	-5.7 (7.5)	-1.5 (3.1)
Age			
11-12	-1.5 (5.3)	-5.4 (6.6)	-1.2 (2.9)
13	-1.6 (5.3)	-7.0 (7.8)	-1.9 (3.1)
14	-1.4 (4.5)	-4.9 (6.2)	-1.1 (2.3)
Level of education			
Basic pre-vocational training	-1.7 (5.9)	-7.2 (8.1)	-1.9 (3.3)
Theoretical pre-vocational training	-1.5 (4.8)	-5.3 (5.0)	-1.2 (2.0)
General secondary education	-1.2 (4.4)	-4.6 (6.0)	-1.1 (2.5)
Country of origin			
The Netherlands	-2.2 (3.8)	-5.1 (6.7)	-0.8 (2.4)
Surinam	-1.6 (4.2)	-6.2 (7.2)	-1.6 (3.0)
Dutch Antilles or Aruba	-1.6 (6.9)	-8.6 (9.1)	-2.5 (3.3)
Morocco	-2.1 (6.1)	-6.7 (6.5)	-1.7 (3.0)
Turkey	-0.4 (5.3)	-7.5 (7.0)	-2.4 (3.0)
Other	-0.6 (6.1)	-4.7 (5.5)	-1.3 (2.5)
Body image			
Far too thin	-1.3 (5.8)	-1.8 (8.5)	0.3 (4.4)
Too thin	-1.0 (4.1)	-2.5 (4.7)	-0.3 (2.0)
Not thin / not fat	-1.7 (5.3)	-5.2 (5.9)	-1.2 (2.4)
Too fat	-1.5 (5.8)	-9.0 (8.3)	-2.5 (3.4)
Far too fat	-1.5 (3.3)	-8.8 (9.9)	-2.5 (3.7)
Weight status (measured)			
Not overweight	-1.8 (4.9)	-3.5 (5.0)	-0.5 (2.1)
Overweight	-1.1 (6.0)	-11.4 (8.0)	-3.6 (3.2)
Time interval			
0-2 month	0.1 (4.5)	-5.2 (7.1)	-1.5 (2.7)
3 month	-1.7 (4.5)	-5.6 (6.5)	-1.3 (2.6)
4 month or more	-2.0 (5.7)	-6.5 (7.3)	-1.6 (3.1)

Table 5.5 Results of hierarchal regression analyses with measured BMI as dependent variable

	R^2	B	95% CI of B	ß	P value
Block 1	0.59				
(Constant)		2.58	1.02-4.14		0.001
BMIs		0.94	0.87-1.00	0.77	0.000
Time interval		0.06	-0.15-0.28	0.02	0.550
Block 2	0.62				
(Constant)		2.19	0.43-3.95		0.015
BMIs		0.91	0.84-0.98	0.74	0.000
Time interval		0.06	-0.16-0.27	0.02	0.593
Gender		0.24	-0.26-0.74	0.03	0.338
Surinam origin		0.69	-0.10-1.47	0.06	0.085
Dutch Antillean / Aruban origin		1.36	0.46-2.28	0.10	0.003
Moroccan origin		0.93	0.11-1.74	0.07	0.026
Turkish origin		1.53	0.76-2.30	0.13	0.000
Other origin		0.45	-0.32-1.22	0.04	0.250
Age		0.29	-0.13-0.71	0.04	0.177
Theoretical pre-vocational training		-0.49	-1.20-0.22	-0.04	0.174
General secondary education		-0.66	-1.24--0.08	-0.07	0.025
Block 3	0.68				
(Constant)		1.20	-0.43-2.83		0.149
BMIs		0.68	0.60-0.76	0.56	0.000
Time interval		-0.02	-0.21-0.18	-0.00	0.874
Gender		0.02	-0.44-0.48	0.00	0.934
Surinam origin		1.06	0.33-1.78	0.09	0.004
Dutch Antillean / Aruban origin		1.66	0.83-2.50	0.12	0.000
Moroccan origin		1.50	0.74-2.26	0.12	0.000
Turkish origin		1.91	1.19-2.62	0.16	0.000
Other origin		0.73	0.02-1.44	0.06	0.044
Age		0.23	-0.16-0.61	0.03	0.252
Theoretical pre-vocational training		-0.49	-1.14-0.16	-0.04	0.139
General secondary education		-0.67	-1.20--0.14	-0.07	0.014
Body image		1.72	1.36-2.07	0.31	0.000

Regression analyses showed that measured BMI was significantly associated with self-reported BMI but also - in order of largest effect size - with body image, country of origin and level of education (Table 5.5). Underestimation of BMI was higher in pupils with a larger body image, pupils from non Dutch origin and pupils from lower levels of education. The time interval between self-report and measurement was not associated with the level of under-reporting. Effect modifications that contributed substantially to the model were not found.

The constant and regression coefficients that were found in the regression analyses (see Table 5.5) were used to calculate a corrected BMI. Two correction formulas were used, one based on BMIs and socio-demographic characteristics alone (explaining 62% of the variance) and the second based on BMIs, socio-demographic characteristics and body image as well (explaining 68% of the variance).

Applying the correction formulas to the total population of 12-13 year olds in Rotterdam yields a prevalence of overweight of 19.2%, when corrected for socio-demographic variables only, or 22.8% when corrected for body image as well, while non-adjusted self-reported data would have resulted in an estimate of 12.5%. Mean BMIs in the total population were adjusted with 1.0 kg/m^2 and 1.3 kg/m^2 respectively, resulting in a mean corrected BMIs of 19.7 kg/m^2 and 19.9 kg/m^2 respectively as compared to a mean non-corrected BMIs of 18.7 kg/m^2.

Sensitivity and specificity for overweight of self-reported BMI were 49% and 96% respectively. Correction of the self-reported BMI for socio-demographic variables resulted in an increase of the sensitivity to 70% and a decrease of the specificity to 89%. Additional correction for body image resulted in a sensitivity and specificity of 77% and 85% respectively.

Based on these estimated prevalences, the positive and negative predictive values of the corrected self-report data have been calculated. For both methods of correcting BMI the positive predictive value was 60% and the negative predictive value 93%.

Discussion

Self-reported data on height and weight of the 12-13 year olds in this study result in a considerable underestimation of BMI and prevalence of overweight. The under-reporting of BMI was higher in pupils who regarded themselves as more fat, were from non Dutch origin and in lower levels of education. A more accurate estimate of BMI was possible when self-reported data were corrected for socio-demographic variables and body image. The sensitivity improved to a great extent when self-reported data were corrected, but still about one fourth of all overweight subjects would not be identified after applying the correction equation.

Underestimation of weight and BMI is found in all of the reviewed studies on validity of self-reported data, although in four of these studies only for females (Brooks-Gunn et al., 1987; Crawley and Portides, 1995; Davis and Gergen, 1994; Strauss, 1999). The degree of underestimation in our study (6.1 kg for weight and 1.5 kg/m^2 for BMI) seems to be larger than in the reviewed studies. Only two studies (Himes et al., 2005; Shannon et al., 1991) report an underestimation of weight greater than 2.5 kg (in both cases for girls only) and only three studies (Brener et al., 2003; Giacchi et al., 1998; Goodman et al., 2000) report an underestimation of BMI with more than 1 kg/m^2 (in two cases only for girls).

There are several possible explanations for the higher level of under-reporting of weight and BMI in our study. In our study pupils knew their self-report data would be treated confidentially and under- or over-reporting would not be noticed either by the school nurse who did the actual measurement or anyone else. In six of the reviewed studies on validity of self-reports, subjects knew self-report would be followed by actual measurement (Brener et al., 2003; Farré Rovira et al., 2002; Himes et al., 2005; Strauss, 1999), which sometimes even took place in the same room (Goodman et al., 2000; Himes and Faricy, 2001). In a confidential situation, like in our study, it is possible that adolescents are more likely to under-report their weight.

Another possible explanation for the high level of under-reporting is that the percentage of overweight subjects in our study population was relatively large (33%). Since overweight subjects appeared to have the high levels of under-reporting BMI this might also contribute to the explanation of our results. The percentage of overweight subjects is not reported in every reviewed study on validity of self-reported height and weight, but in two there are comparable percentages of overweight subjects (34-44% and 33%) and also relatively high under-reporting of weight (1.5 kg and 1.6 kg for boys and 2.1 and 3.5 for girls) (Hauck et al., 1995; Himes et al., 2005).

Still other explanations are the young age of our study population, coinciding with a period of rapid growth, and the relatively low education level.

The time interval between administering self-report and actual measurement in our study was relatively large (three month on average). The difference of 1.5 cm in self-reported and measured height might very well be explained by this time-interval. The under-reporting of weight might also be partly attributed to this time-interval. Our analyses showed that the

length of the time interval was not associated with BMI. But BMI is not expected to change as rapidly over time as height and weight.

Our study also showed that many 12-13 year olds were unable to provide self-reported height and weight data (30% of the total population 12-13 year olds). Himes and Faricy (2001) also report high percentages of missing data in 12 and 13 year olds in the NHANES III study (41% and 25% respectively), Hauck et al. (1995) report 16-17% missing data in a sample of Indian-Americans and Shannon et al. (1991) report 10% implausible or missing data in a sample of sixth graders.

In conclusion, as far as Dutch 12-13 year old pupils in the Rotterdam area are able to provide self-report data on their height and weight, these self-reported data result in a considerable underestimation of BMI and prevalence of overweight. By adjusting self-reported BMI for socio-demographic variables and body image a more accurate estimate of the prevalence of overweight can be given. The usefulness of self-reported height and weight data for the identification of overweight individuals or for use in studies into determinants or correlates of overweight remains questionable. Even if self-report data are adjusted, the sensitivity and specificity do not seem to be sufficient for these purposes.

Applying correction formulas to self-reported height and weight data results in a more accurate estimate of overweight in a society, but since overweight is most probably strongly associated with under-reporting of BMI and prevalence of overweight might differ in different populations or change over time, the usefulness of the correction formulas of the present or any study is limited. Actual measurements in a sub sample will always be needed to draw up new correction formulas for adjustment of self-reported height and weight data in a new sample.

References

Brener ND, Mcmanus T, Galuska DA, Lowry R, Wechsler H (2003), Reliability and validity of self-reported height and weight among high school students. *J Adolesc Health* 32: 281-287.

Brooks-Gunn J, Warren MP, Rosso J, Gargiulo J (1987), Validity of self-report measures of girls' pubertal status. *Child Dev* 58: 829-41.

Brug J, Oenema A, Campbell M (2003), Past, present, and future of computer-tailored nutrition education. *Am J Clin Nutr* 77: 1028S-1034.

Cole TJ, Bellizzi MC, Flegal KM, Dietz WH (2000), Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* 320: 1240-3.

Crawley HF, Portides G (1995), Self-reported versus measured height, weight and body mass index amongst 16-17 year old British teenagers. *Int J Obes Relat Metab Disord* 19: 579-84.

Davis H, Gergen PJ (1994), The weights and heights of Mexican-American adolescents: the accuracy of self-reports. *Am J Public Health* 84: 459-62.

Deckelbaum RJ, Williams CL (2001), Childhood Obesity: The Health Issue. *Obes Res* 9: 239S-243.

Farré Rovira R, Frasquet Pons I, Martínez Martínez MI, Romá Sánchez R (2002), Self-Reported versus Measured Height, Weight and Body Mass Index in Spanish Mediterranean Teenagers: Effects of Gender, Age and Weight on Perceptual Measures of Body Image. *Ann Nutri Metab* 46: 68-72.

Fortenberry JD (1992), Reliability of adolescents' reports of height and weight. *J Adolesc Health* 13: 114-7

Giacchi M, Mattei R, Rossi S (1998), Correction of the self-reported BMI in a teenage population. *Int J Obes Relat Metab Disord* 22: 673-7.

Goodman E, Hinden BR, Khandelwal S (2000), Accuracy of Teen and Parental Reports of Obesity and Body Mass Index. *Pediatrics* 106: 52-58.

Hauck FR, White L, Cao G, Woolf N, Strauss K (1995), Inaccuracy of self-reported weights and heights among American Indian adolescents. *Ann Epidemiol* 5: 386-392.

Himes JH, Faricy A (2001), Validity and reliability of self-reported stature and weight of US adolescents. *Am J Hum Biol* 13: 255-60.

Himes JH, Hannan P, Wall M, Neumark-Sztainer D (2005), Factors Associated with Errors in Self-reports of Stature, Weight, and Body Mass Index in Minnesota Adolescents. *Ann Epidemiol* 15: 272-278.

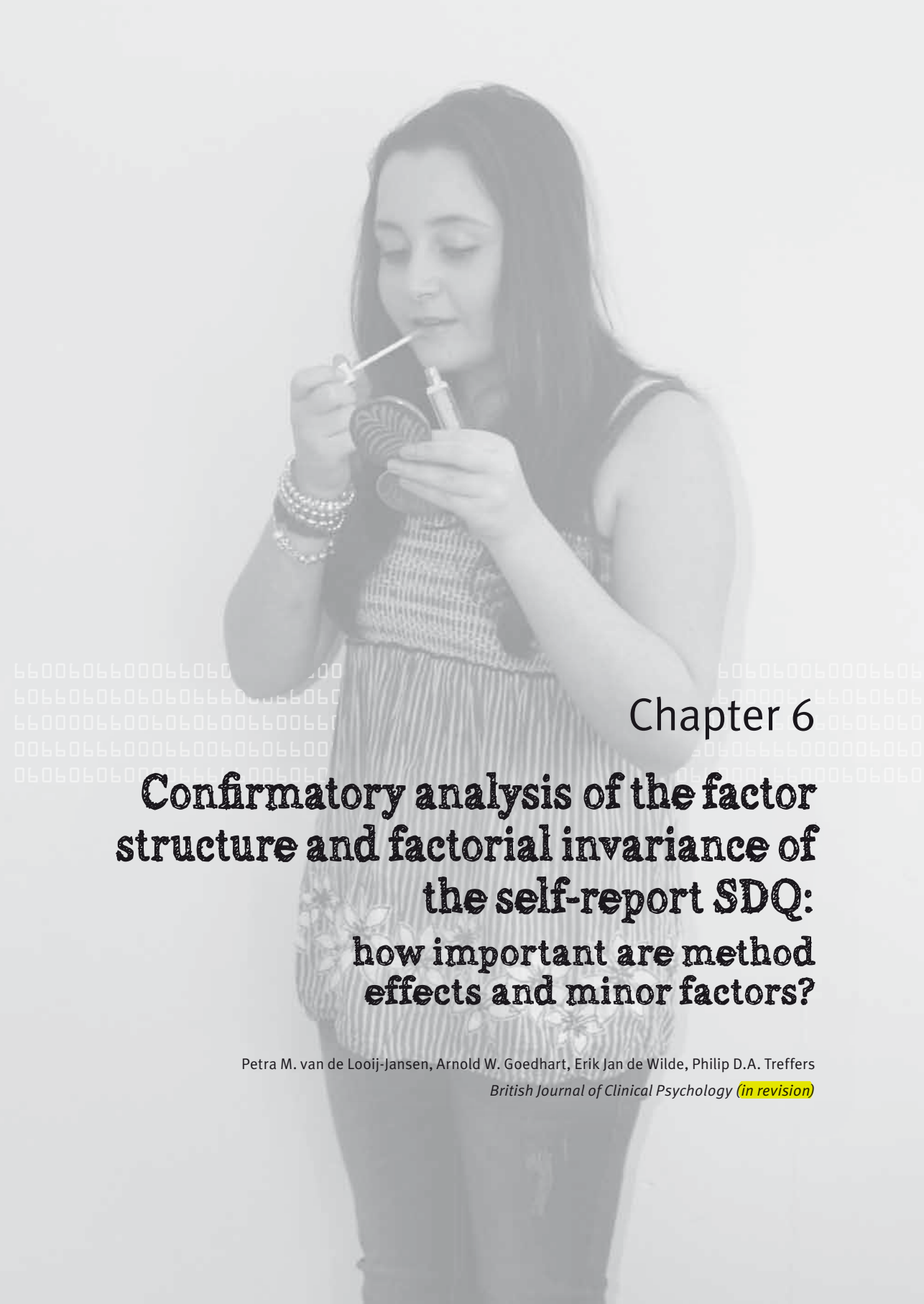
Hirasing RA, Fredriks AM, van Buuren S, Verloove-Vanhorick SP, Wit JM (2001), Toegenomen prevalentie van overgewicht en obesitas bij Nederlandse kinderen en signalering daarvan aan de hand van internationale normen en nieuwe referentiediagrammen. *Ned Tijdschr Geneeskd* 145: 1303-8.

Lobstein T, Frelut ML (2003), Prevalence of overweight among children in Europe. *Obes Rev* 4: 195-200

Shannon B, Smiciklas-Wright H, Wang MQ (1991), Inaccuracies in self-reported weights and heights of a sample of sixth-grade children. *J Am Diet Assoc* 91: 675-8.

Strauss RS (1999), Comparison of measured and self-reported weight and height in a cross-sectional sample of young adolescents. *Int J Obes Relat Metab Disord* 23: 904-8.

Tienboon P, Wahlqvist ML, Rutishauser IH (1992), Self-reported weight and height in adolescents and their parents. *J Adolesc Health* 13: 528-32.



Chapter 6

Confirmatory analysis of the factor structure and factorial invariance of the self-report SDQ: how important are method effects and minor factors?

Petra M. van de Looij-Jansen, Arnold W. Goedhart, Erik Jan de Wilde, Philip D.A. Treffers

British Journal of Clinical Psychology (in revision)

Abstract

Objectives: **The objective of the present study was to examine the factor structure of the self-report Strengths and Difficulties Questionnaire, paying special attention to the number of factors and to negative effects of reverse-worded items and minor factors within the subscales on model fit. Furthermore, measurement invariance across gender, age, level of education and ethnic background was investigated.**

Design: **Data were obtained from the Rotterdam Youth Monitor, a community-based health surveillance system.**

Methods: **The sample consisted of 11,881 pupils of 11-16 years old. Next to the original five-factor model, a factor model with the number of factors based on parallel analysis and scree test was investigated. Confirmatory factor analysis for ordered-categorical measures was applied to examine the goodness-of-fit and measurement invariance of the specified factor models.**

Results: **Best fit of the data was found for the original 5-factor model (i.e., emotional symptoms, conduct problems, hyperactivity-inattention, peer problems and prosocial behaviour) and an alternative 4-factor model, after allowing reverse-worded items to cross-load on the prosocial behaviour factor, and adding error-correlations to the base-models. The 4-factor model differed in that the items of emotional symptoms and peer problems made up one factor. The final 5- and 4-factor models were invariant across gender, age, level of education and ethnic background.**

Conclusions: **While support was found for both models, the final 5-factor model is theoretically more plausible and gained additional support as the original scales emotional problems and peer problems showed significant different relations with gender, educational level and ethnic background.**

Introduction

Early identification and prevention of mental health problems in youth is a major issue. Today, the self-report Strengths and Difficulties Questionnaire (SDQ) is a frequently used screening instrument for mental health problems among adolescents (Goodman et al., 1998; Marzocchi et al., 2004; Obel et al., 2004; Woerner et al., 2004). It consists of five scales of five items each, covering the psychosocial areas of emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and prosocial behaviour. To reduce the effects of response bias (e.g. acquiescence), the items are phrased both positively and negatively, with 10 items reflecting positive attributes ('strengths', including five reverse-worded items of the problem scales) and 15 reflecting negative attributes ('difficulties'). The self-report SDQ, that is applicable to youth from 11 to 16 years of age (Goodman, 1997), can be used as a screening instrument, as part of a clinical assessment, as a treatment-outcome measure, and as a research tool (Goodman, 2001). European studies of the self-report SDQ reported good concurrent and discriminant validity, including strong correlations with the corresponding scales of the Youth Self Report (Klasen et al., 2000; Koskelainen et al., 2000; Muris et al., 2003; Van Widenfelt et al., 2003).

However, several studies on the self-report SDQ have found low internal consistencies (α< .60) for the peer problems and/or conduct problems subscales (Capron et al., 2007; Goodman, 2001; Koskelainen et al., 2000; Koskelainen et al., 2001; Lundh et al., 2008; Mellor, 2004; Muris et al., 2004; Muris et al., 2003; Rønning et al., 2004; Ruchkin et al., 2007; Van Widenfelt et al., 2003). In addition, while most studies using exploratory factor analysis (EFA) with a forced 5-factor structure found support for the original scale structure of the self-report SDQ (Goodman, 2001; Koskelainen et al., 2001; Muris et al., 2004), a number of studies found also support for a 3-factor model (Koskelainen et al., 2001; Percy et al., 2008; Ruchkin et al., 2008) and Muris et al. (2004) found support for a 4-factor model. The studies using confirmatory factor analysis (CFA) reported different results. Three studies found a moderate to good fit to the data for the original 5-factor model (Ruchkin et al., 2008; Ruchkin et al., 2007; Van Roy et al., 2008), two found a questionable fit (Percy et al., 2008; Rønning et al., 2004). Finally, two CFA studies found a comparable fit for 3-factor models (Percy et al., 2008; Ruchkin et al., 2008).

Two highly valued features of the SDQ may be responsible for the relatively low alpha coefficients of some of the subscales and inconsistent results of factor-analytic studies. First, the use of reverse-worded items may be problematic as they may more accurately measure a different construct. For example, Rodebaugh, Woods, and Heimberg (2007) found that reverse-worded social anxiety items were strongly related with extraversion, suggesting that these items primarily assess extraversion. Therefore, the reverse-worded items may negatively influence internal consistencies, because of lower correlations with the remaining items of problem scales, and may cause problems with the factor structure. Second, the small number of items used to cover the rather broad areas of psychosocial adjustment problems may be

problematic as items may represent related but different subareas, leading to less homogeneous scales. For example, the conduct problems scale contains items indicating aggression and items indicating delinquency. The subareas within scales are usually labeled 'minor factors'. This characteristic of the item-set may contribute to low internal consistencies of the scales and indistinctness of the factor structure.

In his analysis of theory and utility of psychological tests and measures for specific purposes, McFall (2005) calls attention to the promising results of inventories, i.e. measures based on 'sampling relatively independent, minimally overlapping items, each contributing uniquely to the probabilistic prediction of the target outcome'. Items of inventories generally show less internal consistency and may not fit neatly into factor analytic solutions. The SDQ-scales have some resemblances with inventories, especially the scales conduct problems and peer problems as each represent a broad problem area, e.g. in the YSR (Achenbach, 1991) the problem area of conduct problems is represented by the subscales delinquent behaviour and aggressive behaviour, the problem area of peer problems is represented by the subscales withdrawn and social problems. As good test-retest reliabilities and predictive validity have been found for the SDQ scales (e.g. Goodman et al., 2000; Lundh et al., 2008), lower levels of internal consistencies (e.g. α.60) of the scales may be considered acceptable. Nevertheless, the grouping of items into scales, i.e. the factorial validity, remains an important research topic. Not only because inconsistent findings have been reported about the latent structure of the self-report SDQ, but especially because important questions on the factor structure have barely been addressed.

First, the crucial decision on the number of factors to retain has been addressed in three studies of the self-report SDQ. These studies used an eigenvalue criterion, i.e. eigenvalue ›1.0 in Koskelainen et al. (2001) and Muris et al. (2004), and eigenvalue ›1.5 in Ruchkin et al. (2008). However, there is increasing consensus among statisticians that two statistically based procedures, parallel analysis (PA) and Velicer's Minimum Average Partial (MAP) test, are superior to the eigenvalue criterion and typically yield optimal solutions to the number of components problem (e.g. O'Connor, 2000). Because Zwick and Velicer (1986) found that PA is generally more accurate if the number of variables is rather small (e.g. 25), we decided to use PA. In addition, we used the scree test to decide on the optimal number of factors and compared the goodness-of-fit of the original 5-factor model with the model based on the number of factors indicated by PA and scree test. The use of multiple criteria to decide on the number of factors is recommended by Fabrigar et al. (1999).

Second, the negative influence of reverse-worded items on the goodness-of-fit of factor models has been examined by considering the influence of reverse-wording as a 'method effect'. Percy et al. (2008) modelled this 'method effect' by adding seven correlations between error terms of reverse-coded items to the model. Van Roy et al. (2008) added a 'positive construal method factor' consisting of the five reverse-worded items to the model. Both studies found a better fit for the modified models. However, the meaning of a 'positive construal method factor' is questionable because the positively worded items of the prosocial

behaviour factor also reflect positive construal, but were not included in the method factor. In the present study we tested the hypothesis that the responses to reverse-worded items reflect both the problem factor and a method effect associated with positive wording. Based on this hypothesis, a significant improvement in model fit was expected after adding secondary loadings of the reverse worded items to the prosocial behaviour factor as these loadings represent effects of positive wording. We preferred a substantial, extended prosocial behaviour factor over a separate method factor as (1) previous studies have found support for the view that method effects associated with positive or negative wording of items may be considered a response style (e.g. DiStefano and Motl, 2006) and (2) method factors have often lead to convergence problems (Kenny and Kashy, 1992). The extended prosocial behaviour factor may reflect a 'positive construal' response style, as suggested by Goodman (2001), or, more specifically 'gamma self-favoring bias', a specific kind of social desirability that is made up of agreeableness and dutifulness (see for review: Paulhus and John, 1998), as this factor is made up by items reflecting agreeableness, e.g. the prosocial behaviour items, and dutifulness, e.g. items 7 (obedient), 21 (reflective) and 25 (persistent).

Next to the factorial validity of the SDQ, another important question that has hardly ever been examined is the invariance of the factor structure across different groups defined by demographic variables, for instance gender or ethnic background. This is an important issue as the usefulness of an instrument depends on its applicability across populations. Ruchkin et al. (2008) found factorial invariance across urban-suburban area, Rønning et al. (2004) found invariance across gender but as their factor model showed a bad fit with both boys and girls, this finding adds little to our knowledge. So, factorial invariance across important domains as gender, age, educational level and ethnic background remains to be studied.

The key aim of the present study is therefore to provide a thorough evaluation of the latent structure of the self-report SDQ. The study extends the literature on the SDQ in several ways. First, this study investigates the number of factors using parallel analysis and the scree test to estimate the optimal number of factors and CFA to compare goodness-of-fit indices of the original 5-factor model and the new factor model. Second, this study looks for minor factors (within the factors) with a substantial negative influence on the goodness-of-fit of the models to the data. Third, confirmatory factor analysis for ordered-categorical measures (CFA-OCM) is used in conducting large sample examinations of factor models, including factor models with minor factors added and/or with secondary loadings of all reverse-worded items on the prosocial factor. Fourth, we examine factorial invariance across gender, age, level of education and ethnic background.

Methods

Study design and procedure

Data were obtained from the Rotterdam Youth Monitor (RYM), a longitudinal youth health surveillance system carried out by the Municipal Public Health Service. The main aim of the RYM is to monitor the general health, well-being, behaviour, and related factors of youth aged 0-19 years living in and around Rotterdam (The Netherlands) in order to supply information for youth policy at school, neighbourhood and municipality level. The RYM is incorporated into the regular health examinations of the preventive youth healthcare system, which means that RYM data are also used to detect potential individual health risks and problems in order to take necessary preventive measures (including referrals for treatment).

For the present study we used data on first and third grade secondary school pupils (respectively, mainly 12-13 and 14-15 year olds). These pupils completed a RYM questionnaire in the period September 2003 to July 2004. Specially trained school nurses of the Municipal Public Health Service carried out the administration of the questionnaire during regular class hours. About 85% of all secondary schools in Rotterdam participate in the RYM.

All data were gathered within and as part of the government approved routine health examinations of the preventive youth healthcare. Separate informed consent was therefore not requested. The questionnaires were completed on a voluntary basis. Parents receive written information on these questionnaires and are free to object to their child's participation. For research purposes, only anonymous data were used.

Participants

12,795 pupils filled out the RYM questionnaire (response rate 90%; non-response mainly due to illness). Cases with missing data on gender, age, level of education, ethnic background and SDQ items were excluded, as well as pupils of 17 years or older. A total of 11,881 pupils were included into the analysis. Of these pupils 50.0% were boys, 40.7% were 11-13 and 59.3% 14-16 years old, 32.3% followed basic and 26.2% theoretical pre-vocational education, 41.5% general secondary education or pre-university training and 48.2% were non-Dutch pupils (11.7% Surinamese, 3.5% Antillean/Aruban, 8.4% Moroccan, 9.8% Turkish, 3.8% Capeverdian and 11.0% other ethnic origin).

Measures

The Dutch translation of the SDQ was administered, as part of the RYM questionnaire (Goodman, 2001; Van Widenfelt et al., 2003). The forward and backward translation procedure and the final version were established in consultation with Goodman, the original author of the instrument (Van Widenfelt et al., 2005). The SDQ includes 25 items, divided into five scales of five items: the emotional symptoms scale, conduct problem scale, hyperactivity-inattention scale, peer problem scale and prosocial scale. The items are scored on a 3-point scale, with scores of 0 for 'not true', 1 for 'somewhat true' and 2 for 'certainly true' for the 15

items reflecting difficulties and the five items of the prosocial scale reflecting strength; the five items of the problem scales that reflect strengths were reverse-scored.

Statistical analysis

Factor models. The sample was randomly divided into two subsamples. The first, model-derivation subsample (n=3960), was used to (1) determine the optimal number of factors, (2) develop a factor model with the optimal number of factors and (3) identify minor factors. The SPSS program for parallel analysis by O'Connor (2000) was used to determine the optimal number of factors. All other factor analyses were performed with Mplus 3.0 (Muthén and Muthén, 1998-2005). Principal factor analysis with Varimax rotation was used to develop a factor model with the optimal number of factors. Minor factors consist of same factor items that represent specific problems within the problems represented by the factor e.g. attention problems within the hyperactivity factor. Because of shared specific content, correlations between items representing a minor factor are substantially higher than predicted by their factor loadings. Therefore, minor factors can be identified by selecting same factor items that show substantial correlation among their error terms. Modification Indices (MI's) provided by Maximum-Likelihood (ML) estimation, indicate the negative influence of error correlations on model fit. We used MI›40 as criterion to select error correlations of same factor items constituting a minor factor that should be added to the model to improve model fit. The second subsample, the model-validation sample (n=7921), was larger to ensure sufficient numbers of cases in all ethnic subgroups when testing for factorial invariance. Robust mean adjusted weighted least square (WLSM) estimation applied to polychoric correlations was used to examine (1) the goodness-of-fit of the original and refined models and (2) the measurement invariance across gender, age, level of education and ethnic background of the final models.

Testing invariance. Measurement invariance reveals the extent to which the essential properties of a measure are independent of the characteristics of the group of persons being measured. If a factor model shows a good fit to the data, then central properties of the factor model are to be regarded as essential properties of the measure and (consequently) factorial invariance will indicate measurement invariance (Millsap, 2007). The levels of factorial invariance are defined by the properties of the factor model that are invariant. Weak factorial invariance, the lowest level of factorial invariance, is defined as invariance of the factorstructure, i.e. the same sets of items are associated with the same factors. The next level, strong factorial invariance, is defined by invariance of the unstandardized factor loadings of the items and with ordered-categorical measures also invariance of the thresholds of the items (estimated as percentiles of the standard distribution), i.e. strong factorial invariance of ordered-categorical measures also implies comparable distributions of each item in different subgroups, e.g. boys and girls (Millsap and Yun-Tein, 2004; Muthén, 1984). Strong factorial invariance is needed for valid comparisons of test scores across groups, e.g. if the SDQ is used as a screening instrument across groups, then strong factorial invariance must be assumed. Finally, we examined structural invariance, i.e. invariance of factor variances and covariances across

groups. Before testing factorial invariance of factorstructure, factor loadings and factor variances and covariances across groups, well-fitting baseline models were estimated separately for each group (e.g. boys and girls) by adding all error correlations ›0.10 between item pairs of the same factor to the model, before testing for factorial invariance (Byrne, 1998; Byrne et al., 1989; Gregorich, 2006; Millsap and Yun-Tein, 2004). The standard hierarchical test-procedure was used (see Millsap & Yun-Tein, 2004; Vandenberg & Lance, 2000) with delta parameterization (see Muthén & Muthén, 1998-2005). Weak factorial (or configural) invariance, the lowest level of factorial invariance denoting equality of the number of factors across groups, was tested first. If weak factorial invariance has been found, then strong factorial invariance was tested. Finally, if strong factorial invariance was found, we tested structural invariance.

Evaluation of goodness-of-fit. The fit of the models was assessed using practical fit indices, the values of which were roughly evaluated using generally employed cut-off criteria (see Marsh et al., 2004). The indices included the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), the root mean square error of approximation (RMSEA) and the Standardised Root Mean square Residual (SRMR). For the CFI and the TLI, values greater than 0.95 were preferred and values near .90 were considered acceptable. RMSEA values of 0.05 or below were preferred, and values below 0.08 were considered acceptable. With respect to the SRMR, values below 0.08 were viewed as indicative of good fit.

The most frequently used tool for testing the difference between models is the chi-square difference test. Recently, though, some researchers have questioned whether this tool should be used, given the chi-square's sensitivity to trivial deviations of fit (Cheung and Rensvold, 2002). We will follow the recommendation of Vandenberg and Lance (2000) to examine the differences in the practical fit indices and use especially the Cheung and Rensvold (2002) CFI-criteria: changes in CFI (ΔCFI) of 0.01 or less indicate that the invariance hypothesis should not be rejected, when ΔCFI lie between 0.01 and 0.02, differences may exist and definite differences between models exist when ΔCFI is greater than 0.02.

Results

Derivation of Factor Models

Using the SDQ data of the model-derivation sample, six eigenvalues › 1 were found using EFA (3.52, 2.42, 1.87, 1.34, 1.06 and 1.01). The scree test indicated that four factors should be retained. The 95th percentiles of random eigenvalues (the first six were: 1.17, 1.14, 1.12, 1.11, 1.09 and 1.08) listed by the Parallel analysis (PA) also indicated that four factors should be retained as the eigenvalue of the 5th SDQ factor (1.06) is lower than the 95th percentile of the 5th random factor. Examination of the 4-factor solution showed that all items of the scales emotional symptoms and peer problems except item 11 (good friend), had loadings ›0.300 on the same factor, labelled 'emotional and peer problems'. The remaining items showed the highest loadings on their original factors, except item 7 (obedient) showing the highest loading on the factor with the hyperactivity items. However, because the model with a secondary loading of item 7 on the hyperactivity factor didn't show a better fit (ΔCFI=0.004) and because the 4-factor and the original 5-factor model should be nested models to compare their fit to the data, item 7 was grouped with the remaining items representing conduct problems and item 11 with the items representing emotional symptoms and peer problems.

With both the 4- and 5-factor models, MI's ›40 were found for error-correlations between items 2 and 10 (indicating a minor factor 'restlessness' within hyperactivity), items 6 and 23 (indicating a minor factor 'loneliness' within (emotional and) peer problems) and items 9 and 20 (indicating a minor factor 'helping' within prosocial). With the 4-factor model additional error correlations with MI ›40 were found between items 16 and 24 (indicating a minor factor 'nervousness' within emotional and peer problems) and items 14 and 19 (indicating a minor factor 'negative peer relations' within emotional and peer problems).

Cross-validation

The 4- and 5-factor base-models showed bad fit to the data of the cross-validation sample (see Table 6.1), with the original 5-factor model showing a significant better fit (ΔCFI=0.03). Extending the prosocial factor into a 'positive construal factor' by allowing reverse-worded items to cross-load on this factor, resulted in much better fit for both models (ΔCFI≥0.040). Adding the minor factors to the base-models resulted in relatively less improvement in the goodness-of-fit indices. Combining both modifications resulted in final models showing good fit to the data. The fit to the data of the final 5-factor model is only slightly better than the fit of the final 4-factor model (ΔCFI=0.011).

As shown in table 6.2, near zero loadings were found for item 11 on (emotional and) peer problems and extended prosocial behaviour and for item 7 on conduct problems. Consequently, these items have a negative influence on the internal consistency of the problem scales. With the exception of item 19 (bullied), low loadings were found for the peer problem items on the emotional and peer problems factor of the 4-factor model.

Table 6.1 Confirmatory Factor Analyses of the self-Report SDQ: Model Fit Indices using WLSM Estimation Applied to Polychoric Correlations.

Estimation Method/Factor model	χ^2	df	CFI	TFI	RMSEA	SRMR
Baseline 5-factor model	7043.14	265	0.874	0.858	0.057	0.065
Baseline 5-factor model with reverse-worded items added to the prosocial factor $\lambda_{pros,7}$, $\lambda_{pros,11}$, $\lambda_{pros,14}$, $\lambda_{pros,21}$, $\lambda_{pros,25}$ free	4897.24	260	0.914	0.901	0.047	0.055
Baseline 5-factor model with correlated errors added: $\theta_{2,10}$, $\theta_{6,23}$, $\theta_{9,20}$ free	6169.64	262	0.890	0.875	0.053	0.062
Final 5 factor model: Baseline 5-factor model with $\lambda_{pros,7}$, $\lambda_{pros,11}$, $\lambda_{pros,14}$, $\lambda_{pros,21}$, $\lambda_{pros,25}$ and $\theta_{2,10}$, $\theta_{6,23}$, $\theta_{9,20}$ free	4214.46	257	0.927	0.914	0.044	0.052
Baseline 4-factor model	8702.04	269	0.844	0.826	0.063	0.073
Baseline 4-factor model with reverse-worded items added to the prosocial factor $\lambda_{pros,7}$, $\lambda_{pros,11}$, $\lambda_{pros,14}$, $\lambda_{pros,21}$, $\lambda_{pros,25}$ free	5910.35	264	0.895	0.881	0.052	0.060
Baseline 4-factor model with correlated errors added: $\theta_{2,10}$, $\theta_{6,23}$, $\theta_{9,20}$, $\theta_{16,24}$, $\theta_{14,19}$ free	7545.02	264	0.865	0.847	0.059	0.069
Final 4-factor model: Baseline 4-factor model with $\lambda_{pros,7}$, $\lambda_{pros,11}$, $\lambda_{pros,14}$, $\lambda_{pros,21}$, $\lambda_{pros,25}$ and $\theta_{2,10}$, $\theta_{6,23}$, $\theta_{9,20}$, $\theta_{16,24}$, $\theta_{14,19}$ free	4793.46	260	0.916	0.903	0.047	0.056

Note: CFI = comparative fit index; TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation; SRMR = standardised root mean square residual.

WLSM = weighted least square using diagonal weight matrix with robust standard errors and mean-adjusted chi-square statistic

Table 6.2 Standardised factor loadings SDQ final 5- and 4-factor models

		5-factor model					4-factor model			
		Ext. prosoc.	Emot. sympt.	Peer probl.	Hyper-activity	Cond. probl.	Ext. prosoc.	E&P probl.	Hyper-activity	Conduct probl.
1.	Considerate	0.73					0.73			
4.	Shares	0.44					0.44			
9.	Caring	0.57					0.57			
17.	Kind to kids	0.63					0.63			
20.	Helps out	0.51					0.51			
3.	Somatic		0.52					0.52		
8.	Worries		0.62					0.62		
13.	Unhappy		0.74					0.74		
16.	Clingy		0.58					0.58		
24.	Fears		0.55					0.55		
6.	Solitary			0.38				0.26		
11*.	Good friend	-0.07		0.07			-0.09	0.00		
14*.	Popular	-0.39		0.43			-0.41	0.27		
19.	Bullied			0.73				0.51		
23.	Better with adults			0.40				0.30		
2.	Restless				0.58				0.59	
10.	Fidgety				0.59				0.59	
15.	Distractible				0.80				0.81	
21*.	Reflective	-0.25			0.41		-0.26		0.40	
25*.	Persistent	-0.17			0.60		-0.18		0.59	
5.	Tempers					0.60				0.63
7*.	Obedient	-0.27				0.06	-0.26			0.07
12.	Fights					0.49				0.53
18.	Lies					0.66				0.66
22.	Steals					0.46				0.47

Note: Ext. prosoc. = Extended prosocial behaviour; E&P probl. = Emotional and Peer problems.

**: reverse-worded items are reverse-coded.*

Internal consistencies, intercorrelations and correlations with demographic variables of the SDQ-scale

The low internal consistency (<0.60) of the scales conduct problems and peer problems reflect low overlap between the items, as found with inventories (see Table 6.3). Cronbach's alpha of the total problem scale was 0.71. The correlations between the scales were generally small (r<0.30), except the correlation between conduct problems and hyperactivity (r=0.37). The moderate correlations between the extended prosocial scale and the problem scales with positively worded items were due to item overlap, as small correlations were found after removing overlapping items from the extended prosocial behaviour scale (e.g. items 11 and 14 when computing the correlation with peer problems).

Different correlations with gender, educational level and ethnic background were found for emotional symptoms and peer problems (see Table 6.3). As found in previous studies (e.g. Koskelainen et al., 2001), girls reported more emotional problems than boys (Cohen's d=0.55, i.e. medium effect size), but the effect size of the difference between boys and girls on peer problems was negligible. In contrast, educational level and ethnic background were related to peer problems but not to emotional problems. More peer problems were reported by adolescents at the lowest educational level (Cohen's d=0.69 for the difference between highest and lowest educational level), this association was also found in other studies (e.g. Flook et al., 2005). Moroccan and Turkish adolescents reported the highest level of peer problems while Dutch adolescents reported the lowest level (Cohen's d=0.61), a similar difference was found between Belgian and immigrant adolescents (Derluyn et al., 2008). These findings are in favour of the 5-factor model as they underline that emotional symptoms and peer problems are different kind of problems.

The correlations of gender and educational level with the extended prosocial scale (see Table 6.3), lend support for the interpretation of the extended prosocial behaviour scale as indicating both agreeableness and dutifulness because these correlations are in line with the results of previous research showing that girls score higher on these traits (e.g. Paulhus and John, 1998) and that dutifulness is positively associated with the educational level of adolescents (e.g. Noftle and Robins, 2007).

Tests of invariance of SDQ factor models

Using the data of sample 2, the CFA-OCM multiple group analysis showed that the hypotheses of weak factorial invariance of the final 5- and 4-factor models could not be rejected for all covariates as a good fit was consistently found (see Table 6.4).

Table 6.3 Cronbach alpha's, intercorrelations and associations with demographic variables of the original SDQ scales, extended prosocial scale and emotional and peer problems scale

	1	2	3	4	5	6	7
1. Emotional symptoms	(0.63)						
2. Conduct problems	0.20	(0.40)					
3. Hyperactivity	0.29	0.37	(0.71)				
4. Peer problems	0.25	0.20	0.04 ns	(0.37)			
5. Prosocial behaviour	0.07	-0.24	-0.19	-0.12	(0.63)		
6. Extended prosocial behaviour	-0.05	-0.41	-0.44	-0.32	0.83	(0.59)	
7. Emotional and Peer problems	0.83	0.25	0.22	0.74	-0.01 ns	-0.21	(0.61)
Gender	0.26	-0.13	0.01 ns	-0.05	0.29	0.17	0.15
Age	-0.01 ns	0.05	0.05	-0.02 ns	-0.03 ns	-0.02 ns	-0.02 ns
Level of education	-0.03 ns	-0.16	-0.04	-0.28	0.07	0.15	-0.18
Ethnic background	0.06	0.12	0.17	0.25	0.05 ns	0.05 ns	0.11

Note: Cronbach alpha's are between brackets; correlations with gender (boys=1, girls=2) are point-biserial correlations, associations with ethnic background are indicated by coefficient eta, the remaining correlations are product-moment correlations; N=7921; correlations and associations not significant at p<0.001 are marked by ns.

When testing the hypotheses of strong factorial invariance and of structural invariance, we observed values of ΔCFI below 0.01 (indicating that these hypotheses could not be rejected) in most cases, but values between 0.01 and 0.02 (indicating that the hypotheses possibly may be rejected) for the covariate gender. We ran a series of analyses to locate the items that caused ΔCFI>0.01 in testing invariance across gender, using the procedures of Cheung & Rensvold (1999). Item 9 (caring) was found to be non-invariant because of very different distributions with boys and girls, i.e. girls were more likely to report caring. Freeing the parameters of item 9 across gender resulted in ΔCFI<0.01 when testing strong factorial and structural invariance for the final 5-factor model. As no definite differences were found when testing the hypotheses of strong factorial and structural invariance (i.e. ΔCFI<.014), and as freeing the constraints of only one out of the ten items of extended prosocial behaviour resulted in negligible differences (i.e. ΔCFI<0.010), the support found for the hypotheses of strong factorial and structural invariance across gender may be considered as sufficient. In a more conservative approach, one could continue working with the scale under partial measurement invariance across gender (Byrne, 1998). With the final 4-factor model, the criterion of ΔCFI<0.01 was reached after freeing the parameters of items 9, 4 and 20 reflecting prosocial behaviour.

Table 6.4 Fit Indices for Invariance of final 5- and 4-factor models across gender, age, level of education and ethnic background

	5-factor model				
	X^2	df	CFI	TLI	RMSEA
Gender					
1. weak factorial invariance	3860.75	496	0.936	0.922	0.041
2. strong factorial invariance	4652.83	541	0.922	0.913	0.044
3. structural invariance	4542.83	556	0.924	0.918	0.043
Age					
1. weak factorial invariance	4096.65	494	0.924	0.920	0.043
2. strong factorial invariance	4334.47	539	0.930	0.922	0.042
3. structural invariance	4318.82	554	0.931	0.925	0.041
Level of education					
1. weak factorial invariance	4443.37	751	0.931	0.918	0.043
2. strong factorial invariance	4716.78	841	0.928	0.923	0.042
3. structural invariance	4717.89	871	0.929	0.926	0.041
Ethnic background					
1. weak factorial invariance	4720.11	1507	0.936	0.924	0.043
2. strong factorial invariance	5226.94	1732	0.930	0.928	0.041
3. structural invariance	5390.94	1807	0.929	0.918	0.041

Note. Age: 12-13 years N=3239, 14-16 years, N=4682; Gender: boys N=3952, girls N=3969; Educational Level: low N=2532, low-middle N=2038, middle-high N=3351; Ethnic background: Dutch N=4099, Surinamese N=924, Antillean N=288, Moroccan N= 664, Turkish N=784, Cape Verdian N=284.

4-factor model				
χ^2	df	CFI	TLI	RMSEA
4234.06	500	0.929	0.915	0.043
5107.11	547	0.913	0.905	0.046
4989.97	557	0.916	0.909	0.045
4619.27	503	0.924	0.910	0.045
4843.38	550	0.921	0.914	0.044
4870.35	560	0.921	0.915	0.044
4573.48	796	0.930	0.921	0.042
5044.20	838	0.922	0.916	0.044
4960.09	858	0.924	0.920	0.043
4768.47	1493	0.935	0.921	0.043
5450.77	1728	0.926	0.923	0.043
5535.35	1778	0.925	0.924	0.042

Discussion

In this study, we thoroughly evaluated the factor structure of the adolescent self-report SDQ in a multicultural community sample of Dutch adolescents, that was split into a model-derivation and a cross-validation sample. First, we examined the number of factors. Using the data of the model-derivation sample, both parallel analysis and scree test indicated that four factors should be retained. However, the original 5-factor model showed a much better fit to the data of the cross-validation sample than the 4-factor model (with a factor made up of the items of the emotional symptoms and peer problems scales and factors made up by the items of the conduct problems, hyperactivity and prosocial behaviour scales, respectively). The final conclusion on the number of factors is that a 5-factor model should be maintained because of (1) difficulties in interpreting the combined emotional and peer problems factor with four peer problem items showing low loadings (≤ 0.30) and (2) the scales emotional symptoms and peer problems showed, in line with previous studies, very different associations with gender, educational level and ethnic background, indicating that these scales represent substantially different problem areas.

Second, three minor factors were identified using the data of the model derivation sample: 'restlessness' (within the hyperactivity-inattention factor), 'loneliness' (within the peer problems factor), and 'helping' (within the prosocial factor). As the minor factors represent specific problems within the problem area indicated by their factors, the correlations between the error-terms of the item-pairs that make up the minor factors (item-pairs 2 - 10, 6 - 23 and 9 - 20) represent reliable variance and should be added to the model (see Green and Herschberger, 2000). A significant improvement in fit of the model to the data of the cross-validation sample was found after adding the three correlated error-terms. The previous studies of Percy et al. (2008), Van Roy et al. (2008) and Rønning et al. (2004) also added the correlated error between items 2 and 10 to their models, but didn't present a systematic analysis of minor factors with a significant negative influence on model fit. Van Roy et al. (2008) added the error correlation between items 18 (lies) and 19 (bullied) to their model, but didn't explain why that error correlation may indicate reliable variance.

Third, a substantial improvement in fit of the original 5-factor model to the data of the cross-validation sample was obtained after adding secondary loadings of the reverse-worded items on the prosocial behaviour factor. This finding lend support to our hypothesis that the responses to the reverse-worded items should be regarded as reflecting not only the presence of a specific problem, but also a response style associated with the positive wording of these items and the items of prosocial behaviour, e.g. gamma self-deception.

Fourth, factorial invariance across gender, age, level of education and ethnic background was found for the final five-factor model. However, it must be noted that the hypothesis of strong factorial and structural invariance across gender was nor clearly acceptable nor definitely refuted. But, as freeing the invariance constraints for only 1 of the 10 items of the extended prosocial behaviour factor, in this case item 9 (caring), resulted in acceptable dif-

ferences in levels of fit, we regard the hypotheses as not refuted. These findings indicate that the items of the self-report SDQ are not differently interpreted across those subgroups. So, when differences between scale-scores are reported across gender, age, level of education or ethnic background, these are more likely real differences, rather than differences in the interpretation of the SDQ items. Hence, valid comparisons across these groups can be made. This is an important finding as the self-report SDQ is being used in many large epidemiological studies.

In this study and in the studies by Percy et al. (2008) and Rønning et al. (2004) a bad fit was found for the original (not modified) 5-factor model, while other studies found a good fit (Ruchkin et al., 2007; Ruchkin et al., 2008; Van Roy et al. 2008). As we found factorial invariance across demographic variables, we couldn't imagine any explanation in terms of sample differences for these different results. Therefore, we decided to replicate the three previous studies that reported on the programs and estimation methods used (i.e. Percy et al., 2008; Rønning et al., 2004; Van Roy et al., 2008), using comparable programs and methods. Much to our surprise, we found comparable levels of fit as reported in each of these studies, indicating that the differences may not be caused by differences between samples, but by differences between programs and methods, especially between Mplus WLSM estimation (indicating a rather bad fit) and LISREL DWLS estimation (as used by Van Roy et al., 2008, indicating a good fit). While a simulation study by Hu (posted on the Mplus website) showed that the Mplus approach worked well, a simulation study comparing the two approaches is needed.

One of the strengths of our study is the large sample size. This meant that we were able to compose a model-derivation and a model-validation subsample. Furthermore, the large sample size enabled us to study measurement invariance across different levels of education and different ethnic groups. A limitation of our study is that the results are limited to the self-report SDQ. As the SDQ also consists of a parent and teacher form, it is recommended to investigate the factorial validity and factorial invariance of the parent- and teacher version of the SDQ in the same way. Also, clinical samples should be included. We further recommend investigating the concurrent and discriminant validity of the extended prosocial behaviour factor of the final 5-factor models.

In conclusion, we found good fit to our sample data with the final 5-factor models although some adjustments are recommended. In practice, this means that the same problems scales can be used as specified in the original SDQ and use of the extended prosocial behaviour scale should be preferred. We further conclude that the self-report SDQ is applicable for use in different populations, provided that the extended prosocial behaviour scale is used, as measurement invariance was found for gender, age, level of education, and ethnic background.

References

Achenbach TM (1991), *Manual for the Child Behavior Checklist/4-18 and1991 Profile*. Burlington (VT): University of Vermont, Department of Psychiatry.

Byrne BM (1998), *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic applications and programs*. Mahwah, NJ: Lawrence Erlbaum.

Byrne BM, Shavelon RJ, Muthén BO (1989), Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychol Bull* 105: 456-466.

Capron C, Therond C, Duyme M (2007), Psychometric properties of the French version of the self-report and teacher Strengths and Difficulties Questionnaire (SDQ). *Eur J Psychol Assess* 23: 79-88.

Cheung GW, Rensvold RB (2002), Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equat Model* 9: 233-255.

Derluyn I, Broekaart E, Schuyten G (2008), Emotional and behavioural problems in migrant adolescents in Belgium. *Eur Child Adolesc Psychiatry* 17: 54-62.

DiStefano C, Motl RW (2006), Further investigating method effects associated with negatively worded items on self-report surveys. *Struct Equat Model* 13: 440-464.

Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999), Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods* 4: 272-299.

Flook L, Repetti RL, Ullman JB (2005), Classroom social experiences as predictors of academic performance. *Dev Psychol* 41: 319-327.

Goodman R (1997), The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 38: 581-586.

Goodman R (2001), Psychometric properties of the Strengths and Difficulties Questionnaire. *J Am Acad Child Adolesc Psychiatry* 40: 1337-1345.

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

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

Green SB, Herschberger SL (2000), Correlated errors in true score models and their effect on coefficient alpha. *Struct Equat Model* 7: 251-70.

Gregorich SE (2006), Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Med Care* 44: S78-94.

Kenny DA, Kashy DA (1992), Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychol Bull* 112: 165-172.

Klasen H, Woerner W, Wolke D, Meyer R, Overmeyer S, Kaschnitz W, Rothenberger A, Goodman R (2000), Comparing the German versions of the Strengths and Difficulties Questionnaire (SDQ-Deu) and the Child Behavior Checklist. *Eur Child Adolesc Psychiatry* 9: 271-276.

Koskelainen M, Sourander A, Kaljonen A (2000), The Strengths and Difficulties Questionnaire among Finnish school-aged children and adolescents. *Eur Child Adolesc Psychiatry* 9: 277-284.

Koskelainen M, Sourander A, Vauras M (2001), Self-reported strengths and difficulties in a community sample of Finnish adolescents. *Eur Child Adolesc Psychiatry* 10: 180-185.

Lundh LG, Wangby-Lundh M, Bjarehed J (2008), Self-reported emotional and behavioral problems in Swedish 14 to 15-year-old adolescents: A study with the self-report version of the Strengths and Difficulties Questionnaire. *Scand J Psychol* 49: 523-532.

Marsh HW, Hau KT, Wen Z (2004), In search of golden rules: Comment on hypothesis testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct Equat Model* 11: 320-41.

Marzocchi GM, Capron C, Di Pietro M, Duran Tauleria E, Duyme M, Frigerio A, Gaspar MF, Hamilton H, Pithon G, Simoes A, Therond C (2004), The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *Eur Child Adolesc Psychiatry* 13: II40-6.

Mc Fall RM (2005), Theory and utility - Key themes in evidence-based assessment: Comment on the special section. *Psychol Assess* 17: 312-323.

Mellor D (2004), Furthering the use of the strengths and difficulties questionnaire: reliability with younger child respondents. *Psychol Assess* 16: 396-401.

Millsap RE (2007), Invariance in measurement and prediction revisited. *Psychometrika* 72: 461-473.

Millsap RE, Yun-Tein J (2004), Assessing factorial invariance in ordered-categorical measures. *Multivariate Behav Res* 39: 479-515.

Muris P, Meesters C, Eijkelenboom A, Vincken M (2004), The self-report version of the Strengths and Difficulties Questionnaire: its psychometric properties in 8- to 13-year-old non-clinical children. *Br J Clin Psychol* 43: 437-448.

Muris P, Meesters C, van den Berg F (2003), 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* 12: 1-8.

Muthén BO (1984), A general structure equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika* 49: 115-132.

Muthén LK, Muthén BO (1998-2005), *Mplus User's Guide* Los Angeles, C.A.: Muthén & Muthén.

Noftle EE, Robins RW (2007), Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *J Pers Soc Psychol* 93: 116-130.

O'Connor BP (2000), SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav Res Methods, Instrum Comput* 32: 396-402.

Obel C, Heiervang E, Rodriguez A, Heyerdahl S, Smedje H, Sourander A, Guethmundsson OO, Clench-Aas J, Christensen E, Heian F, Mathiesen KS, Magnusson P, Njarethvik U, Koskelainen M, Ronning JA, Stormark KM, Olsen J (2004), The Strengths and Difficulties Questionnaire in the Nordic countries. *Eur Child Adolesc Psychiatry* 13: II32-9.

Paulhus DL, John OP (1998), Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *J Pers* 66: 1025-1060.

Percy A, McCrystal P, Higgins K (2008), Confirmatory factor analysis of the Adolescent Self-Report Strengths and Difficulties Questionnaire. *Eur J Psychol Assess* 24: 43-48.

Rodebaugh TL, Woods CM, Heimberg RG (2007), The reverse of social anxiety is not always the opposite: The reverse-scored items of the social interaction anxiety scale do not belong. *Behav Ther* 38: 192-206.

Rønning JA, Handegaard BH, Sourander A, Morch WT (2004), The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *Eur Child Adolesc Psychiatry* 13: 73-82.

Ruchkin V, Jones S, Vermeiren R, Schwab-Stone M (2008), The Strengths and Difficulties Questionnaire: the self-report version in American urban and suburban youth. *Psychol Assess* 20: 175-182.

Ruchkin V, Koposov R, Schwab-Stone M (2007), The Strength and Difficulties Questionnaire: scale validation with Russian adolescents. *J Clin Psychol* 63: 861-9.

Van Roy B, Veenstra M, Clench-Aas J (2008), Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *J Child Psychol Psychiatry* 49: 1304-1312.

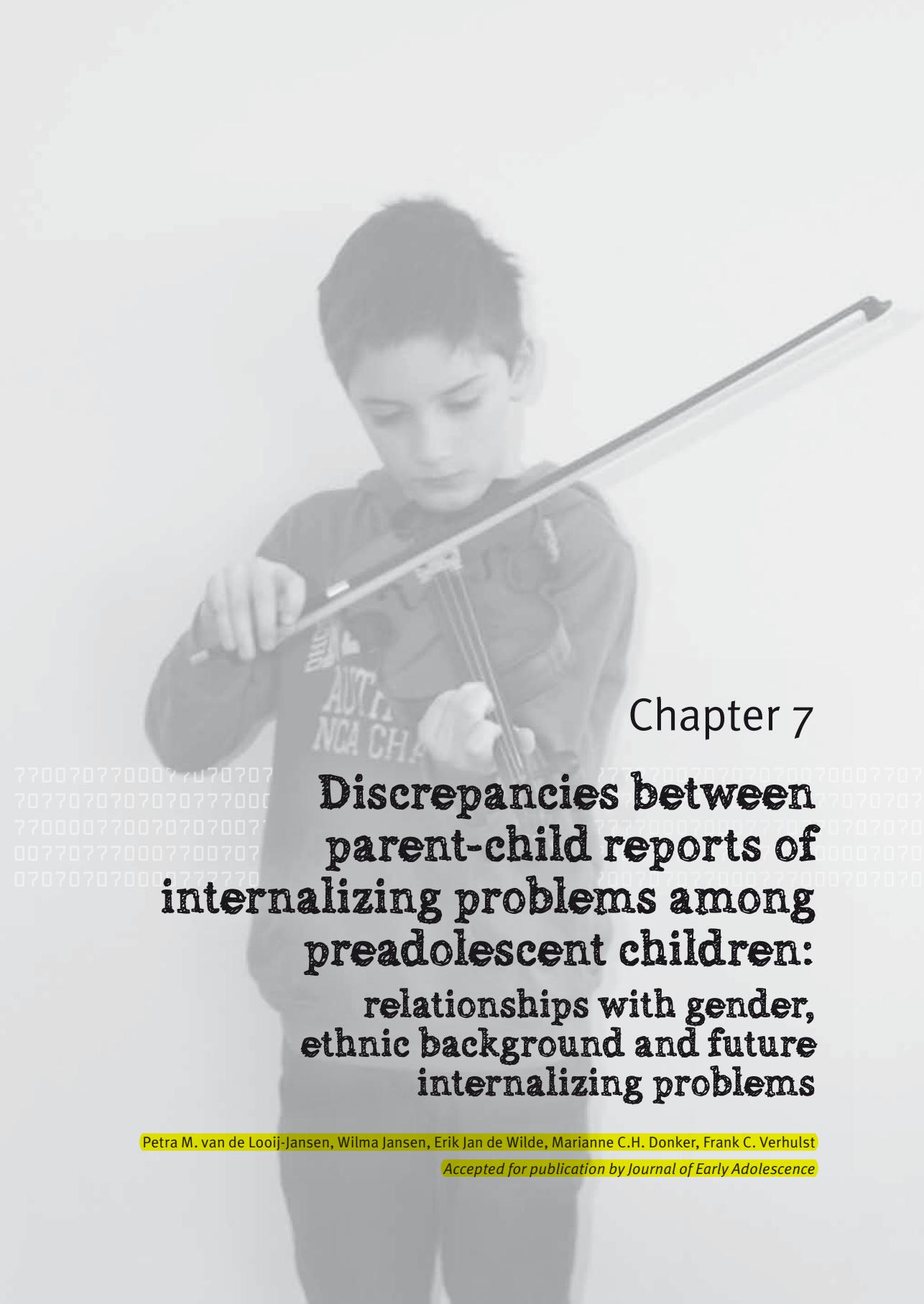
Van Widenfelt BM, Goedhart AW, Treffers PD, Goodman R (2003), Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 12: 281-9.

Van Widenfelt BM, Treffers PD, de Beurs E, Siebelink BM, Koudijs E (2005), Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clin Child Fam Psychol Rev* 8: 135-47.

Vandenberg RJ, Lance CE (2000), A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Res Methods* 3: 4-70.

Woerner W, Fleitlich-Bilyk B, Martinussen R, Fletcher J, Cucchiaro G, Dalgalarrondo P, Lui M, Tannock R (2004), The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. *Eur Child Adolesc Psychiatry* 13: II47-54.

Zwick WR, Velicer WF (1986), Comparison of five rules for determining the number of components to retain. *Psychol Bull* 99: 432-442.



Chapter 7

Discrepancies between parent-child reports of internalizing problems among preadolescent children:

relationships with gender, ethnic background and future internalizing problems

Petra M. van de Looij-Jansen, Wilma Jansen, Erik Jan de Wilde, Marianne C.H. Donker, Frank C. Verhulst

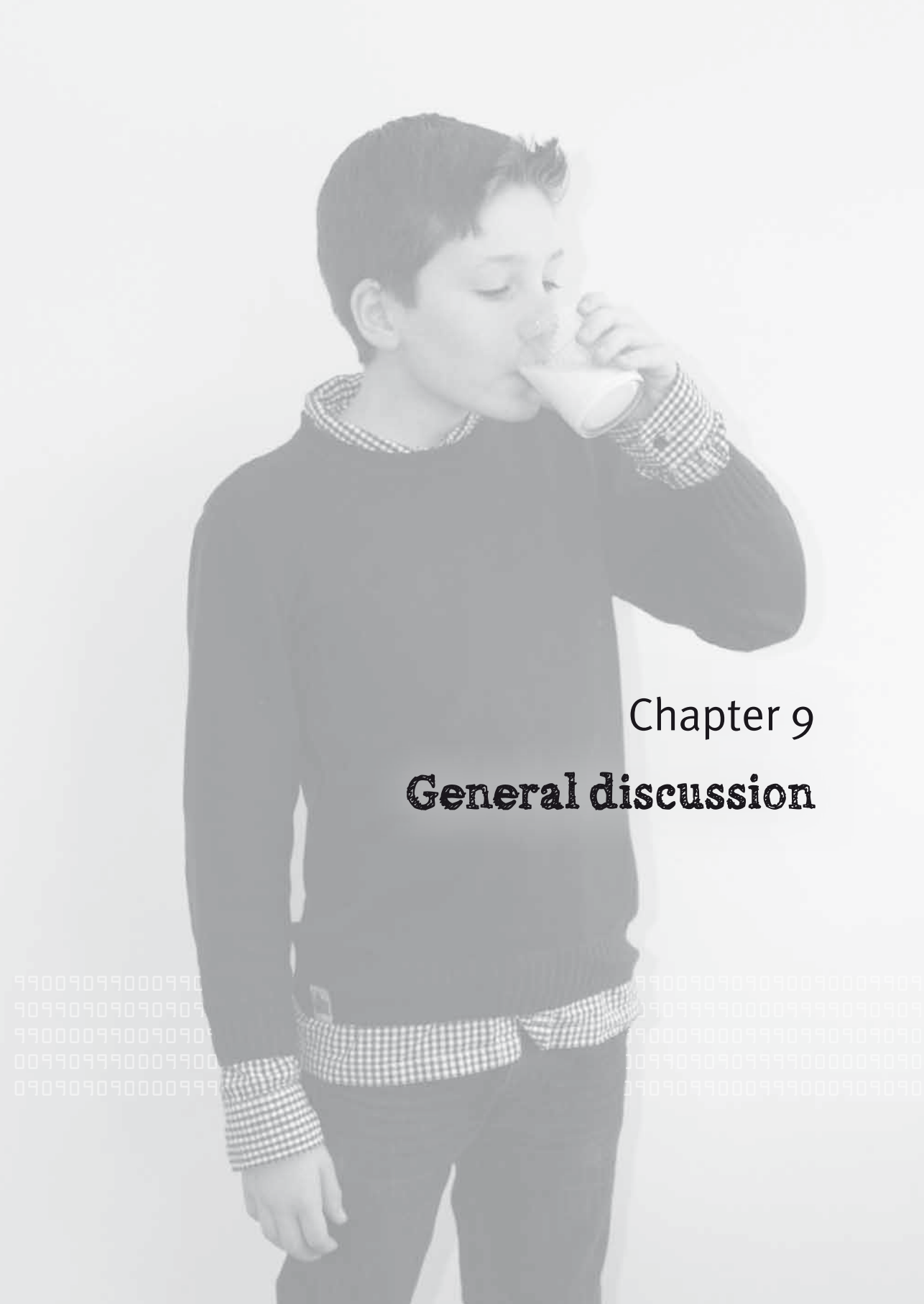
Accepted for publication by Journal of Early Adolescence

Chapter 8

Self-report of suicidal thoughts and behaviour and school nurse evaluations in Dutch high school students

Erik Jan de Wilde, Petra M. van de Looij-Jansen, Judith E.J. Goldschmeding, Christina Hoogeveen

Submitted



Chapter 9

General discussion

So far, this thesis has reported on a number of methodological and validity issues related to the use of self-report questionnaires in the Rotterdam Youth Monitor (RYM). The final chapter discusses the main findings, outlines the strengths and limitations of the various studies, and proposes directions for further research and public health practice.

Main findings

Method of data collection

One general conclusion on the studies presented in this thesis is that different methods of data collection did not produce greatly different results. Whether we varied the anonymity of the questionnaire (anonymous versus confidential), the way in which the responses were provided (paper versus computer), or the season of administration, the differences between groups were small or non-significant. These studies therefore suggest that self-report is a robust method in a youth healthcare setting. However, this general conclusion needs to be refined.

The first study investigated differences in health and health-related behaviour between anonymous and confidential data-collection procedures, an issue never previously examined by studies among adolescents in the Netherlands. As stated, anonymous and confidential data collection produced no significant differences for most health indicators, the exceptions being good perceived health, use of alcohol, and aggressive behaviour (the latter only for boys), which were reported significantly less frequent in the confidential condition.

We therefore concluded that confidential self-report measures make it possible to derive epidemiologically comparable outcomes for most health indicators, and that these are also valuable for detecting individual at-risk children. This conclusion is supported by our finding that pupils who score high on social desirability did not respond differently to most health indicators in either condition. Although we did not experimentally investigate the effect of confidentiality assurances, the minimal differences in health indicators between the two conditions may have been due to the effort we made to assure pupils of the confidentiality of the questionnaire. In confidential data-collection procedures which include sensitive questions, it is important to assure respondents that their data will be kept confidential. A meta-analysis by Singer et al. (1995) indicated that these assurances generally boost overall response rates and item-response rates to the sensitive questions.

The second study investigated differences in health and health-related behaviour between web-based and paper-and-pencil data-collection procedures. Due to their many advantages – such as cost advantages, automatic branching, and inconsistency checking – computerised questionnaires are now being used more and more. However, it is still not clear whether, as some authors suggest, adolescents disclose more sensitive information in computerised questionnaires (Turner et al., 1998; Wang et al., 2005).

While Mangunkusumo et al. (2005) investigated differences in the Netherlands between a web-based and a pencil-and-paper mode of data collection, their questionnaire included mainly non-sensitive topics. Concluding that sensitive issues among youth had not been sufficiently determined, Mangunkusumo recommended that future research should explore any differences produced by the two data-collection methods in answers on sensitive topics (Mangunkusumo, 2006). Our own study investigated a wide variety of topics, including sensitive ones. The main finding is that a web-based RYM questionnaire yielded almost the same results as pencil-and-paper one: there were no differences on topics one would suppose to be sensitive, such as vandalism, stealing, sexual behaviour, or the use of alcohol or marijuana. The only exception was for carrying a weapon, which had a greater number of reports in the web-based condition. Neither were there any differences in perceived level of privacy and confidentiality between the two conditions.

A remark should be made with regard to the implementation of web-based administration in the daily practice of the RYM. Starting in the 2006-2007 school year, we gradually implemented web-based RYM questionnaires at secondary schools (which rose from 31% in 2006-2007, to 64% in 2007-2008, and to 90% in 2008-2009). As not all secondary schools are well-equipped with computer labs, 100% web-based coverage of these questionnaires has not yet been achieved – and probably never will be. As a consequence, mixed modes of data collection are still daily practice with the RYM. The results of our study suggest that this does not affect the early identification of health problems and prevalence rates of health indicators.

The third study on survey methodology covered seasonal variations in self-reported health and health-related behaviour – a second issue that had not previously been investigated among young people in the Netherlands. We found no strong evidence of major seasonal variations either in self-reported health or mental health, or in health-related behaviour. Though the strongest effect was found for alcohol consumption, where scores were highest in May and June, this difference remained far below Cohen's criteria for a small effect size. It therefore seems unnecessary for data collection in youth health surveys to be spread over the year.

Plainly, the results of these three studies do not imply that administration in any setting yield reliable results. A study by Brener et al. (2006), who also included *setting* (school versus home) in their design, concluded that *mode* effects (pencil-and-paper versus computer) were more weakly associated with self-reported health-risking behaviour among high school students than *setting* effects were (school versus home). Setting was associated with 30 of the 55 risk behaviours examined, whereas mode was associated significantly with the reporting of seven of the 55 behaviours.

Recommendations for the administration of the questionnaire are made later in this chapter.

Validity of self-report measures

The four studies that focus on the validity of self-report measures do not provide evidence for a single overall conclusion. We can nonetheless remark that while self-report does not in itself guarantee valid measurement, it may very well provide meaningful predictive information.

The truth of this remark was demonstrated by the first study, which investigated the differences between self-reported height and weight, and measured height and weight. Due to the expense involved in measuring the height and weight of large group of individuals, youth health surveys often use self-report.

In the Netherlands, however, the validity of self-reported height and weight in adolescents has never been evaluated. Our study showed that self-report on height and weight led the prevalence of overweight to be considerably underestimated in three groups of pupils: those who regarded themselves as fatter, those of non-Dutch origin, and those of lower educational levels.

We concluded that, unless the figures are adjusted, self-reported height and weight are an inappropriate method for estimating overweight prevalence in 12 to 13-year-olds. In daily practice, adjustment would mean that actual measurements of weight and height in a subsample will always be needed to compose new correction formulae for adjusting the self-reported heights and weights produced in a new sample.

At an individual level, the use of adjusted self-reported BMI is questionable. For these purposes, neither the sensitivity nor the specificity seems sufficient. The implication of this for preventive youth healthcare is that there is only one way of gaining true insight into the percentage of young people with overweight or underweight: the measurement of height and weight in routine health examinations.

The second study thoroughly examined the factor structure of the adolescent self-report SDQ, paying special attention to the number of factors, to negative effects of reverse-worded items, and to the effect of minor factors within the subscales on model fit. It also examined measurement invariance across gender, age, level of education and ethnic background.

After reverse-worded items had been allowed to cross-load on the prosocial factor, and after error-correlations had been added to the base-models, the best fit of the data for the self-report SDQ was found for the original five-factor model and an alternative four-factor model. We also showed that the refined five- and four-factor models were invariant across gender, age, level of education and ethnic background. This finding indicates that the items of the self-report SDQ had not been interpreted differently across those subgroups. So, when differences are reported across gender, age, level of education or ethnic background, these are more likely to be real differences, not differences in the interpretation of the SDQ items.

The third study investigated how discrepancies between parent-child reports of internalizing problems are related with gender, ethnic background, and future internalizing problems. Parent-child disagreement of internalizing problems varied across ethnic groups, the largest parent-child discrepancies occurring with regard to children from Surinamese/Antillean and

Turkish backgrounds. While more self-reported than parent-reported internalizing problems were found in Surinamese/Antillean children, Turkish children had more parent-reported internalizing problems that than self-reported ones. Disagreement between parents and their preadolescent child significantly contributed to the prediction of self-reported internalizing problems in early adolescence.

The fourth study examined current practice with regard to school nurses' screening and follow-up actions, particularly with regard to the relation between self-reported suicidal thoughts and behaviour, and the judgements and subsequent action of school nurses. The vast majority of invitations for a routine health examination were not followed by any action on the part of the school nurse. Neither did the nurse consider that most students whose self-report led to a 'point of special interest' were actually at risk. Another finding is that the self-report of recent suicidal thoughts added to the predictive value of emotional distress as measured by the SDQ and other self-report conclusions.

Limitations and strengths

Design and population

One characteristic of all the studies described in this thesis is that they were carried out in a school setting: after all, schools are the main points of access for preventive youth healthcare, and thus for the RYM, which is an integral part of it.

Although a main advantage of the school setting is the high response rate (in our samples at secondary schools it was around 90%), non-response by truants and out-of-school youth may have led to selection bias. But although some studies have found a higher prevalence of risk behaviours – such as alcohol use, drug use, sexual intercourse and delinquent behaviour among truants and dropouts (Brener, 1998; Weller, 1997; Swaim, 1997) – we do not believe that, for the purposes of our studies, such factors would have greatly affected our conclusions. One reason is our assumption that the percentage of non-response due to truants and dropouts was distributed equally between anonymous and confidential groups and between the web-based and the pencil-and-paper group.

Even though it does not refer to a limitation of our study, another remark might be made about the school setting. Although the selection bias inherent to it might cause the prevalence of health indicators to be underestimated, schools are excellently suited to youth health monitoring, as they are important points of access not only for preventive youth healthcare monitoring, but also for health-promotion programmes. Under the RYM, schools receive their own health profile every two years (based on RYM results), and are supported with the development of a school health policy. This is a favourable contrast with household surveys, which face a variety of problems, such as low response rates, possible influence by parents, and the selection bias that may be caused if ethnic minorities are under-represented (Vollebergh et al., 2006).

A strength of the studies described in this thesis is the practice-based setting inherent to preventive youth healthcare and the RYM, which gave rise to several methodological and validity issues. The advantage of a practice-based approach is precisely its basis in daily practice rather than in experimental or laboratory situations. Partly as a consequence of this, any results are likely to be implemented directly – and possibly more successfully – in daily practice. The approach also provides an opportunity to increase the research-mindedness of those involved in the daily practice of preventive youth healthcare, for instance by involving health professionals in research projects from the earliest possible stages.

Other strengths lie not only in the high response rates on the part of secondary schools (85-90%) and their pupils (approximately 90%), but also in the ethnic diversity of the population. The large samples of respondents enabled us to conduct studies that required large numbers of respondents, such as our studies of seasonal variation and of the factorial validity and factorial invariance of the self-report SDQ for which model-derivation and model-validation sub samples were constructed. This ethnic diversity enabled us to study not only the relationships between ethnic background and discrepancies between parent-child reports of emotional problems, but also measurement invariance of the self-report SDQ across pupils from different ethnic backgrounds.

Measures

A first limitation of the two studies on the predictive validity of emotional problems and suicide is the absence of a clinical outcome measure. As a criterion, we had to use other self-reports or judgements by a nurse. This can be explained by the fact that the task of preventive youth healthcare in a stepped-care approach is not to diagnose, but to identify health problems early, and, if necessary, to refer to more specialised service. A remark with regard to the judgements of school nurses as outcome measure is that Vogels et al. (2008) found serious inter-professional variation in the identification of emotional and behavioural problems. This remains as an important subject for both research and professionalisation.

A second issue concerns the questionnaires used. In general, when young people's health is monitored, it is recommended that similar versions of questionnaires are used. Also, when multiple sources are used, it is recommended to use a scale or questionnaire that has separate versions for multiple informants (parent, child, teacher) (e.g. Verhulst and Van der Ende, 2008). In our study it was not possible to follow these recommendations, since there is no instrument for monitoring the health of preadolescent children (i.e. aged 9-10 in our study) into early adolescence. The commonest instruments with separate versions for multiple informants are the self-report SDQ and the Youth Self Report (YSR), which are suitable for people aged 11-16 (SDQ) or 11-18 (YSR). To study discrepancies in parent-child report of emotional problems we therefore had to rely on instruments in which the parent questionnaire covered a wider range of emotional problems than the child questionnaire.

Indications for future research

Given the commonness of self-report questionnaires in youth and adult research, it is remarkable how few Dutch studies have examined the methodological issues related to the use of these questionnaires with young people. It is therefore necessary to answer several questions related to their use in youth health surveys, which would generally require empirical methodological and validation studies.

More specifically, future research should address issues related to the issues listed below. These fall into three distinct categories: 1) the validity of (RYM) measures across cultures, 2) the use of the RYM across all types of education, and 3) young people's concerns about confidentiality and privacy.

Validity of (RYM) measures across cultures

Within the setting of preventive youth healthcare, the SDQ is widely regarded as a very promising questionnaire for the early identification of psychosocial problems. Originally developed by Goodman in the UK (Goodman, 2001), it was translated into Dutch by van Widenfelt et al. (van Widenfelt et al., 2003; van Widenfelt et al., 2005). Although valid translations of questionnaires are important, they are only a beginning. As even a well-translated questionnaire item may involve a concept that has different meanings in different cultures, subscales developed on the basis of factor analyses in one population may not be valid factors in another (Knox and Echeveria, 2009). As the usefulness of an instrument depends on its applicability across populations, this is a highly relevant issue.

However, a general shortcoming of most validation studies in the Netherlands is that the psychometric properties of questionnaires have been tested only in Dutch populations. In Rotterdam – the second largest city in the Netherlands – 61.1% of 0-20 year olds are not ethnically Dutch. The largest immigrant groups are Surinamese, Turkish and Moroccan (Statistics Netherlands, 2008). The increasing ethnic diversity of young populations indicates a growing need for greater certainty about the cross-cultural validity of various instruments, such as those recommended for use in preventive youth healthcare. By measuring invariance across different groups (gender, age, level of education, and ethnicity) for the self-report SDQ, this thesis has therefore taken an important first step towards gaining such certainty. As the SDQ also consists of a parent and teacher form, we recommend that a future study uses a similar technique to measure invariance in the parent and teacher versions of the SDQ.

More research is also needed on the cross-cultural validity of other instruments used in the RYM questionnaires – an issue that has already been picked up in the 'Academische Werkplaats Diversiteit in Jeugdbeleid in de Stadsregio Rotterdam-Rijnmond' (*www.samendwars.nl*).

The use of the RYM across all types of education

Until now, the RYM has been carried out only in regular secondary schools. Schools for special education have not yet been involved, as it is assumed that the RYM questionnaires are too long and difficult for this specific population. Question comprehension is indeed known to be one of the basic cognitive processes that can cause error – and thus affect validity – in the question-answering process (Bowling, 2005; Brener et al., 2003). But schools for special education in Rotterdam increasingly desire insight into the health and health-related behaviour of their populations, for instance with regard to alcohol consumption or physical activity.

Because it is also important for the RYM itself to monitor the health of all young people – including those in special education – we recommended research is conducted to establish whether it is necessary to adapt the questions or the length of the questionnaire for this specific group. And given that pupils in special education generally have poorer reading skills, it would also be reasonable to investigate other methods of data collection, such as audio self-administered questionnaires (ASAQ), in which a portable cassette player plays a recording of the questions, and respondents mark their answers on a paper answer sheet. An alternative method is audio computer-assisted self-administered interviewing (ACASI), in which the computer displays the question on screen and simultaneously plays a digitised recording of it to the respondent, typically via earphones (Tourangeau et al., 2000).

Young people's concerns about confidentiality and privacy

Another issue relates to concerns about confidentiality and privacy. Although paper-and-pencil and web-based data-collection procedures did not seem to differ with regard to perceived level of privacy and confidentiality, a substantial percentage of pupils reported confidentiality and privacy concerns. Further research should establish which factors play a role in pupils' (and parents') concerns about privacy and confidentiality.

In this respect, it would also be interesting to investigate whether the current political and media debate in the Netherlands on the use of Electronic Child Recording in preventive youth healthcare has any influence on pupils' concerns about privacy and confidentiality. One possible hypothesis is that the current debate is detrimental to the early identification of problems – including psychosocial ones – as young people may distrust assurances of confidentiality, and thus wonder what will happen with the information they provide. Such distrust might then cause them to withdraw from participation in the RYM, or to underreport any health problems or health-related behaviour.

Another issue related to confidentiality and privacy involves the extent to which respondents should be reminded that their privacy is protected – for it is possible to remind them too much. If too much stress is laid on protecting privacy, respondents can become suspicious. Assurances can thus become counterproductive (Lensvelt-Mulders, 2008).

Implications for public health practice

As a means of collecting data at individual and collective levels alike, the RYM is a unique instrument for preventive youth healthcare and youth policy. However, because an overall summary of respondents' answers ('points of special interest') is used as an input for the routine health examinations by the school nurse, a confidential data-collection procedure is also needed. When administering a questionnaire, it is important to guarantee sufficient confidentiality and privacy, as any lack of perceived confidentiality or privacy may cause response bias (Brener, 2003). Special attention should therefore be paid to ensuring that the procedure for collecting data in a classroom is – and is seen to be – confidential.

Observations made during the administration of the questionnaire suggest that there is scope for improving data-collection procedures by school nurses and assistants. To make data-collection more professional, we therefore make two recommendations.

First, communication with schools should pay greater attention to the purpose of the questionnaire and of the procedures required for confidential data collection in the classroom. School directors and teachers should be convinced of the importance of guaranteeing sufficient confidentiality during questionnaire administration. To collect data confidentially, an examination setting is preferred. If, alternatively, computer labs are used, partitions between the computers would maximise pupils' sense of privacy.

The second recommendation is that, before the questionnaire is filled in, the school nurse should give it a full verbal introduction, and pupils should be assured of its confidentiality. For these purposes, a data collection-protocol should be developed, and school nurses and assistants should receive annual training in confidential questionnaire administration. To guarantee the quality of assessments, youth healthcare professionals should develop and implement proper guidelines.

Despite such measures, it will be difficult to ensure that conditions for collecting data in the classroom are always consistent. Every effort should nonetheless be made to guarantee that the data-collection procedure maximises confidentiality.

Questionnaire content and indications for future research

The RYM questionnaire for secondary school pupils consists of a basic part intended to identify problems at an individual level, and a flexible part covering topics which often vary from year to year, and are sometimes requested by local organisations. One example is the questions asked within the framework of 'Communities that Care', a community strategy for promoting healthy youth development and reducing levels of youth drug use, violence, delinquency, teenage pregnancy, and school drop-out (Arthur et al., 2002; Pollard et al., 1999).

Overall, the RYM questionnaire thus contains a variety of topics, including 'sensitive' ones – those which, according to Tourangeau and Yan (2007), respondents may perceive as intru-

sive because they raise fears about the potential repercussions of disclosing the information, or because they trigger concerns about social desirability. The extent to which a topic is perceived as sensitive often differs across countries, ethnic groups and age groups. For instance, it is less threatening to ask about smoking marijuana in the Netherlands than it is in the United States (Lensvelt-Mulders, 2008). In adults questions about income and voting are regarded as sensitive topics (Tourangeau et al., 2000).

Because we do not really know which questions are sensitive for young people in the Netherlands, we recommend that research is conducted to identify the specific questions they regard as 'sensitive' or 'too personal'. Such research should also examine differences across gender, age, and ethnic background.

Another issue related to content of questionnaires involves the possible negative impact of answering 'sensitive' questions. While it is often suggested – for instance by parents or schools involved in the RYM – that young people may find it upsetting to fill in such a personal questionnaire, the distress they actually feel when answering personal survey questions has seldom been examined.

The results of the few studies that have been conducted on this are inconclusive. A study by Langhinrichsen-Rohling et al. (2006) showed that 4.4% of the study sample were often upset while completing a survey. Although suicidal behaviour, a history of physical abuse, and/or a history of sexual abuse all emerged as significant predictors, they accounted for only 6.6% of the variance in upset ratings.

These results contrast with those of a randomised controlled study by Gould et al. (2005) which examined the iatrogenic risk of youth-suicide screening programmes. This showed that experimental groups (who used a questionnaire with suicide questions) did not differ from control groups (who used a questionnaire without suicide questions) with regard to distress levels 1) immediately after the first survey or 2) two days later. Neither did rates of depressive feelings differ. Students who were asked questions on suicide were no more likely to report suicidal ideation after the survey than those who were not. It was concluded that suicide screening had no iatrogenic effects.

Such inconclusiveness leads us to recommend that future research investigates the impact on adolescents' well-being of answering the RYM questionnaire.

Overall, we believe that greater attention should be paid to the quality of self-report questionnaires and data-collection procedures. Evidence-based guidelines should be developed, especially with regard to questionnaire length. For example, it should be decided which length is acceptable for pupils at pre-vocational education. Answers on questionnaire content should also be answered: for instance, is it acceptable to ask first grade pupils questions on drug use or sexual intercourse? We also recommend that a quality mark should be developed for survey questionnaires.

Both of these issues might be further elaborated in the *Lokale en Nationale Monitor Jeugdgezondheid* project (*www.monitorgezondheid.nl*), which intends to develop standard questions for health surveys and registration in preventive youth healthcare.

In summary, research on the validity and reliability of questionnaires in a practical youth healthcare setting is relatively new, especially in the Netherlands. While self-report seems to be promising, its quality can be improved only after further research. Better self-report will affect the quality of youth healthcare and prevention, and would contribute to the further professionalisation of the youth healthcare sector.

References

Arthur MW, Hawkins JD, Pollard JA, Catalano RF, Baglioni AJ, Jr. (2002), Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors. The Communities That Care Youth Survey. *Eval Rev* 26: 575-601.

Bowling A (2005), Mode of questionnaire administration can have serious effects on data quality. *J Publ Health (Oxf)* 27: 281-91.

Brener ND, Billy JOG, Grady WR (2003), Assessment of factors affecting the validity of self-reported health-risk behavior among adolescents: Evidence from the scientific literature. *J Adolesc Health* 33: 436-457.

Brener ND, Eaton DK, Kann L, Grunbaum J, Gross LA, Kyle TM, Ross JG (2006), The association of survey setting and mode with self-reported health risk behaviors among high school students. *Public Opin Q* 70: 354-374.

Goodman R (2001), Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry* 40: 1337-1345.

Gould MS, Marrocco FA, Kleinman M, Thomas JG, Mostkoff K, Cote J, Davies M (2005), Evaluating iatrogenic risk of youth suicide screening programs: a randomized controlled trial. *Jama* 293: 1635-43.

Knox S, Echeveria D (2009), Methodological issues related to longitudinal epidemiological assessment of developmental trajectories in children. *J Epidemiol Community Health* 63 (Suppl I): i1-i3.

Langhinrichsen-Rohling J, Arata C, O'Brien N, Bowers D, Klibert J (2006), Sensitive research with adolescents: just how upsetting are self-report surveys anyway? *Violence and Victims* 21: 425-444.

Lensvelt-Mulders GJLM (2008), Surveying Sensitive Topics. In: *International Handbook of Survey Methodology*, de Leeuw ED, Hox JJ, Dillman DA, eds. New York: Lawrence Erlbaum Associates, pp 461-478.

Mangunkusumo R (2006), *E-health4Uth: integrating Monitoring and Prevention in Youth Health Care*. Thesis.

Mangunkusumo R, Moorman P, Berg van den-Ruiter de A, Lei van der J, Koning de H, Raat H (2005), Internet-administered adolescent questionnaires compared with a paper version in a randomized study. *J Adolesc Health* 36: 70-76.

Pollard JA, Hawkins JD, Arthur MW (1999), 'Risk and protection: are both necessary to understand diverse behavioral outcomes in adolescence?' *Social Work Research* 23: 145-158.

Singer E, von Thurn D, Miller E (1995), Confidentiality assurances and response: A quantitative review of the experimental literature. *Publ Opin Q* 59: 257-266.

Tourangeau R, Rips L, Rasinski K (2000), *The psychology of survey response* Cambridge, UK: Cambridge University Press.

Tourangeau R, Yan T (2007), Sensitive questions in surveys. *Psychol Bull* 133: 859-83.

Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL (1998), Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science* 280: 867-73.

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

Van Widenfelt BM, Treffers PD, de Beurs E, Siebelink BM, Koudijs E (2005), Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clin Child Fam Psychol Rev* 8: 135-47.

Verhulst FC, Van der Ende J (2008), Using rating scales in a clinical context. In: *Rutter's Child and Adolescent Psychiatry*, 5th edition, Rutter M, Bishop D, Pine D, Scott S, Stevenson J, Taylor E, Thapar A, eds.: Blackwell Publishing.

Vogels AGC, Jacobusse GW, Hoekstra F, Brugman E, Crone MR, Reijneveld SA (2008), Identification of children with psychosocial problems differed between preventive child health care professionals. *J Clin Epidemiol* 61: 1144-1151.

Vollebergh WAM, Van Dorsselaer S, Monshouwer K, Verdurmen J, Van der Ende J, Ter Bogt T (2006), Mental health problems in early adolescents in the Netherlands. Differences between school and household surveys. *Soc Psychiatry Psychiatr Epidemiol*: 156-163.

Wang YC, Lee CM, Lew-Ting CY, Hsiao CK, Chen DR, Chen WJ (2005), Survey of substance use among high school students in Taipei: web-based questionnaire versus paper-and-pencil questionnaire. *J Adolesc Health* 37: 289-95.

Summary

This thesis reports about a number of methodological and validity issues related to the use of self-report questionnaires among youth in the setting of preventive youth healthcare. Seven specific research questions are derived from the Rotterdam Youth Monitor (RYM), a longitudinal youth health surveillance system integrated in preventive youth healthcare in Rotterdam and the surrounding municipalities. The studies described in this thesis are divided into two parts. The *first* part concerns the extent to which the reports on these questionnaires depend on the circumstances of administration. These issues include different methods of data collection (anonymous versus confidential self-report questionnaire; web-based versus paper-and-pencil self-report questionnaire) and seasonal variation in self-reported health and health-related behaviour. The *second* part concerns the validity of measures used in the RYM: the validity of self-reported length and weight, the factorial validity of the self-report Strengths and Difficulties Questionnaire in various groups of young people, the predictive validity of discrepancies in parent-reported and child-reported emotional problems, and the predictive validity of emotional problems (including suicide) for the various actions taken by school nurses.

Chapter 1 is the introduction to this thesis and describes the context of the Rotterdam Youth Monitor, different perspectives on validity and the research questions.

Chapter 2 describes the comparison of responses on indicators of (mental) health and health-related behaviour in two different groups of data collection: anonymous and confidential. 39 fifth-grade classes with 704 mainly 16-17 year old adolescents were randomly allocated to the anonymous (n=351) or confidential condition (n=353). Pupils in the confidential condition were told that completed questionnaires would "never be shown to teachers, parents, police or anyone else, except to the school nurse". Pupils in the anonymous condition were told that no one would be able to identify them personally. Anonymous and confidential data collection produced no significant differences for most health indicators, the exceptions being good perceived health, use of alcohol, and aggressive behaviour (the latter only for boys), which were reported significantly less frequent in the confidential condition. It appeared that pupils with high scores on social desirability did not respond differently at most health indicators in the two conditions. For most health indicators, epidemiologically comparable outcomes may be derived from confidential self-report measures, which are also valuable for detecting individual at-risk children.

Chapter 3 focuses on differences in responses related to (mental) health and behaviour between a web-based (web) and paper-and-pencil (p&p) data collection procedure. Privacy and confidentiality concerns were investigated as a possible source of response bias. Of the five participating schools all third-grade classes (mainly 14-15 year olds pupils) were ran-

domly assigned to either the Internet condition (n=271) or the paper-and-pencil condition (n=261). Significant but small differences were found for the Strengths and Difficulties subscales 'emotional symptoms' (p&p>web) and 'prosocial behaviour' (p&p>web), and carrying a weapon (web>p&p). Perceived level of privacy and confidentiality did not differ between the two modes. The findings suggest that in a controlled school setting, web-based administration of health indicators yields almost the same results as paper-and-pencil administration.

Chapter 4 examines whether self-reported (mental) health and health-related behaviour varies by season. A total of 33,171 first and third grade pupils (mainly 12-15 year olds) completed a questionnaire of the Rotterdam Youth Monitor in the period 2003-2006. Significant but small seasonal differences were found for psychological well-being, self-esteem, the subscales 'conduct problems', 'hyperactivity', 'peer problems', and the total problem score of the Strengths and Difficulties Questionnaire, having breakfast, sporting activities, watching television, and alcohol consumption. Gender-by-season interactions were hardly found. The strongest effect was found for alcohol consumption, showing the highest scores in May-June. However, all differences remained far below Cohen's criteria for a small effect size. We found no strong evidence for major seasonal variation in self-reported (mental) health and health-related behaviour among adolescents in the Netherlands. It therefore seems unnecessary for data collection in youth health surveys to be spread over the year.

The purpose of **chapter 5** was to investigate the validity of self-reported height and weight in 12- to 13-year-olds, to identify sociodemographic correlates and to explore whether correction factors can be developed to estimate the prevalence of overweight in youth. 5525 12-13-year old pupils in the Rotterdam area filled in a confidential questionnaire on health topics, including their height and weight. In a sub-sample of 499 pupils both self-reported and measured height and weight were available. Self-reported data led to a considerable underestimation of Body Mass Index (BMI) and consequently the prevalence of overweight. Underestimation was higher in pupils who regarded themselves as more fat, were of non-Dutch origin and in lower educational levels. Self-reported height and weight appeared to be inappropriate to estimate the overweight prevalence in 12-to-13-year-olds, unless the figures were adjusted. Using adjusted self-reported BMI on an individual level is questionable. Actual measurements of height and weight are necessary to draw up valid correction formulas in new samples.

Chapter 6 examines the factor structure of the self-report Strengths and Difficulties Questionnaire (SDQ), paying special attention to the number of factors, the negative effects of reverse-worded items and the effect of minor factors within the subscales on model fit. It also examines measurement invariance across gender, age, level of education and ethnic background. The sample consisted of 11,881 first and third grade pupils (mainly 12-15 year olds). Confirmatory factor analysis for ordered-categorical measures was applied to exam-

ine the goodness-of-fit and measurement invariance of the specified factor models. After reverse-worded items had been allowed to cross-load on the prosocial factor, and after error-correlations had been added to the base-models, the best fit of the data for the self-report SDQ was found for the original five-factor model and an alternative four-factor model. Our study further showed that the refined five- and four-factor models were invariant across gender, age, level of education and ethnicity. This finding indicates that the items of the self-report SDQ are not differently interpreted across those subgroups. So, when differences are reported across gender, age, level of education or ethnicity, these are more likely to be real differences, not differences in the interpretation of the SDQ items.

In **chapter 7** we investigated how discrepancies between parent-child reports of internalizing problems are related with gender, ethnic background, and future internalizing problems. No significant differences in discrepancy scores were found between boys and girls. Parent-child disagreement of internalizing problems varied across ethnic groups, with the largest parent-child discrepancies occurring within children from Surinamese/Antillean (children reporting more internalizing problems than their parents) and Turkish background (parents reporting more internalizing problems than their children). Disagreement between parents and their preadolescent child significantly contributed to the prediction of self-reported internalizing problems in early adolescence. No support was found for a moderation effect of gender or ethnic background on the association of discrepancy scores and future internalizing problems. For the early identification and treatment of internalizing problems it is recommended to include both parent and child self-reports as part of routine health examinations in the setting of preventive youth healthcare. Also, in a society that increasingly becomes multicultural, healthcare professionals should be more aware of cultural discrepancies in parent and child reported internalizing problems.

Chapter 8 describes the current practice with regard to school nurses' screening and follow-up actions, particularly with regard to the relation between self-reported suicidal thoughts and behaviour, and the judgements and subsequent action of school nurses. For this purpose, data of 3692 first grade pupils were analysed. The vast majority of invitations for a routine health examination were not followed by any action on the part of the school nurse. Neither did the nurse consider that most students whose self-report led to a 'point of special interest' were actually at risk. Another finding is that the self-report of recent suicidal thoughts added to the predictive value of emotional distress as measured by the SDQ and other self-report conclusions.

Finally, **chapter 9** contains a general discussion of the findings presented in this thesis, and addresses directions for future research and implications for public health practice.

A general conclusion of the various studies in this thesis is that the effect of variations in method of data collection is limited. Whether we varied the anonymity of the question-

naire (anonymous or confidential), the mode of administration of the responses (paper or computer), or the season of administration, differences between groups were small or non-significant. The four studies that focussed on the validity of self-report measures did not provide one general conclusion.

For future research it is recommended to conduct studies on the cross-cultural validity of self-report measures used in youth health surveys, to investigate if questions have to be adjusted for pupils at schools for special secondary education and to explore if other methods of data collections are more suitable for this specific group. Other suggestions are to investigate which factors are related to pupils' confidentiality and privacy concerns and to investigate the impact of filling in the RYM questionnaire on the well-being of adolescents.

Samenvatting

Dit proefschrift gaat in op aspecten die te maken hebben met de methode en de validiteit van vragenlijsten die door jongeren worden ingevuld. De zeven onderzoeksvragen die in dit proefschrift aan de orde komen, vloeiden voort uit de dagelijkse praktijk van de Jeugdmonitor Rotterdam, een instrument waarmee de gezondheid van kinderen en jongeren tot en met 18 jaar wordt gemonitord. Het proefschrift bestaat uit twee delen. Het eerste deel gaat in op de mate waarin zelfrapportage wordt beïnvloed door de omstandigheden waarin de vragenlijst wordt afgenomen. Hieronder vallen verschillende methoden van dataverzameling (anoniem versus vertrouwelijk; internet/digitaal versus schriftelijke afname) en seizoensinvloeden in zelfrapportage van gezondheid en aan gezondheid gerelateerd gedrag. Het tweede deel gaat in op de validiteit van vragen die gebruikt worden in de Jeugdmonitor Rotterdam, namelijk de validiteit van zelfgerapporteerde gegevens over lengte en gewicht, de factoriële validiteit van de zelfrapportage versie van de Strengths and Difficulties Questionnaire (SDQ), de predictieve validiteit van het al dan niet overeenkomen van zelfgerapporteerde gegevens over emotionele problemen tussen ouder en kind, en de predictieve validiteit van emotionele problemen voor wat betreft de acties die jeugdverpleegkundigen naar aanleiding hiervan ondernemen.

Hoofdstuk 1 bevat de introductie van dit proefschrift en geeft een beschrijving van de Jeugdmonitor Rotterdam en de onderzoeksvragen.

Hoofdstuk 2 beschrijft een onderzoek waarin de antwoorden op vragen over gezondheid en aan gezondheid gerelateerd gedrag vergeleken zijn in twee verschillende groepen van dataverzameling: anoniem en vertrouwelijk. 39 vijfde klassen met 704 voornamelijk 16-17 jarigen werden willekeurig verdeeld over een anonieme (n=351) en een vertrouwelijke conditie (n=353). Leerlingen in de vertrouwelijke conditie werd verteld dat ingevulde vragenlijsten nooit gezien zouden kunnen worden door leerkrachten, ouders, politie of iemand anders, met uitzondering van de jeugdverpleegkundige. Leerlingen in de anonieme conditie werd verteld dat niemand hen persoonlijk zou kunnen identificeren. Voor de meeste gezondheidsindicatoren zijn geen significante verschillen gevonden tussen een anonieme en een vertrouwelijke dataverzameling. Uitzonderingen hierop zijn een goede ervaren gezondheid, het gebruik van alcohol, en agressief gedrag (alleen bij jongens), die significant minder vaak werden gerapporteerd in de vertrouwelijke conditie. Jongeren met een hoge score op sociale wenselijkheid bleken in de twee condities niet verschillend te antwoorden op de meeste gezondheidsindicatoren. Voor de meeste gezondheidsindicatoren levert een vertrouwelijk afgenomen vragenlijst vergelijkbare epidemiologische uitkomsten op als een anoniem afgenomen vragenlijst, en deze uitkomsten zijn ook waardevol voor het signaleren van individuele risicojongeren

De focus van **hoofdstuk 3** is een onderzoek naar de verschillen in zelfrapportage van gezondheid en aan gezondheid gerelateerd gedrag tussen een digitale (via het web afgenomen) en een schriftelijke vragenlijst. De mate waarin leerlingen zich zorgen maken over privacy en vertrouwelijkheid zijn onderzocht als mogelijke bronnen van respons bias. Van de vijf deelnemende scholen zijn alle derdeklassers (voornamelijk 14-15 jarigen) willekeurig toegewezen aan een internet conditie (n=271) en een papieren conditie (n=261). Significante, maar kleine verschillen zijn gevonden voor de SDQ subschalen emotionele problemen (papier›web) en prosociaal gedrag (papier›web), en het dragen van een wapen (web›papier). De mate waarin leerlingen zich zorgen maakten over privacy en vertrouwelijkheid verschilde niet tussen de twee condities. De resultaten geven aan dat in een gecontroleerde schoolsetting, een digitale afname van een gezondheidsvragenlijst vergelijkbare antwoorden oplevert als een papieren afname.

Hoofdstuk 4 beschrijft een studie waarin is onderzocht of zelfrapportage van gezondheid en aan gezondheid gerelateerd gedrag verschilt naar seizoen van afname. 33.171 brugklassers en derdeklassers (voornamelijk 12-15 jarigen) vulden een vragenlijst in van de Jeugdmonitor Rotterdam in de periode 2003-2006. Significante, maar kleine verschillen naar seizoen zijn gevonden voor psychisch welbevinden, zelfbeeld, gedragsproblemen (SDQ), hyperactiviteit (SDQ), problemen met leeftijdsgenoten (SDQ), de totale probleemscore op de SDQ, ontbijten, sporten, televisie kijken en de consumptie van alcohol. Seizoensverschillen waren nauwelijks verschillend tussen jongens en meisjes. Het sterkste effect is gevonden voor de consumptie van alcohol, met het hoogste gebruik in mei en juni. Echter, alle verschillen die gevonden zijn waren klein volgens de criteria van Cohen. We vonden geen sterk bewijs voor grote seizoensvariatie in zelfrapportage van gezondheid en aan gezondheid gerelateerd gedrag bij jongeren in Nederland. Het is daarom niet noodzakelijk om de dataverzameling bij onderzoek naar de gezondheid van jongeren te spreiden over het jaar.

Hoofdstuk 5 onderzoekt de validiteit van zelfrapportage van lengte en gewicht bij 12-13 jarigen, bekijkt welke socio-demografische kenmerken hiermee samenhangen en verkent of een correctiefactor ontwikkeld kan worden om de prevalentie van overgewicht te schatten. 5525 brugklassers vulden een vertrouwelijke gezondheidsvragenlijst in, met o.a. vragen over lengte en gewicht. In een subpopulatie van 499 leerlingen waren zowel zelfgerapporteerde als gemeten waarden van lengte en gewicht beschikbaar. De resultaten laten zien dat zelfgerapporteerde data tot een aanzienlijke onderschatting van de Body Mass Index (BMI) leiden en hiermee van het voorkomen van overgewicht. Onderschatting was sterker bij leerlingen die zichzelf te dik voelen, bij leerlingen van niet-Nederlandse herkomst en bij leerlingen van lagere schoolniveaus. Zelfrapportage van lengte en gewicht is ongeschikt om de prevalentie van overgewicht te schatten bij 12-13 jarigen, tenzij de data gecorrigeerd worden. Een gecorrigeerde BMI op basis van zelfrapportage van lengte en gewicht op individueel niveau is twijfelachtig. Actuele metingen van lengte en gewicht zijn nodig om valide correctie formules te verkrijgen in nieuwe steekproeven.

Hoofdstuk 6 gaat in op de factorstructuur van de zelfrapportage versie van de Strengths and Difficulties Questionnaire (SDQ), met speciale aandacht voor het aantal factoren, de negatieve invloed van contra-indicatieve items en het effect van ondergeschikte factoren binnen de subschalen op de model fit. Daarnaast is onderzocht of de factorstructuur invariant is naar geslacht, leeftijd, schoolniveau en etnische herkomst. De steekproef bestond uit 11.881 brugklassers en derdeklassers (voornamelijk 12-15 jarigen). Confirmatieve factoranalyse voor variabelen met een ordinaal meetniveau is toegepast om de goodness-of-fit en invariantie van de gespecificeerde modellen te onderzoeken.

Het toestaan dat de contra-indicatieve items ook op de prosociale factor laadden, en het toevoegen van correlaties tussen de meetfouten aan de basismodellen, resulteerde in een goede fit voor het originele 5-factoren model en een alternatief 4-factoren model. Onze studie liet verder zien dat de aangepaste 5- en 4-factormodellen invariant waren voor geslacht, leeftijd, schoolniveau en etnische herkomst. Dit betekent dat de items van de zelfrapportage versie van de SDQ niet verschillend geïnterpreteerd worden bij deze verschillende subgroepen. Wanneer verschillen naar geslacht, leeftijd, schoolniveau of etnische herkomst gevonden worden, dan duidt dit op werkelijke verschillen en niet op verschillen in de interpretatie van SDQ items.

In **hoofdstuk 7** is onderzocht of discrepanties in de rapportage van emotionele problemen tussen ouder en kind samenhangen met het geslacht van het kind, etnische herkomst en toekomstige emotionele problemen. Discrepanties tussen ouder en kind zijn niet verschillend voor jongens en meisjes. Wel zijn verschillen naar etnische herkomst gevonden, waarbij de grootste discrepanties gevonden zijn voor Surinaamse/Antilliaanse kinderen (kinderen rapporteren meer problemen dan hun ouders) en Turkse kinderen (ouders rapporteren meer problemen dan hun kinderen). Het niet overeenkomen van ouder en kind droeg significant bij aan de voorspelling van zelfgerapporteerde emotionele problemen in de vroege adolescentie. Geen steun is gevonden voor een moderatie effect van geslacht of etnische herkomst op de relatie tusen discrepantie scores en toekomstige emotionele problemen. Voor de vroegtijdige onderkenning en behandeling van emotionele problemen wordt aanbevolen om zowel de rapportage van ouders als kind als onderdeel van het preventieve gezondheidsonderzoek in de jeugdgezondheidszorg mee te nemen. Verder zouden professionals in de gezondheidszorg zich meer bewust moeten zijn van culturele verschillen in discrepanties tussen ouder en kindgerapporteerde emotionele problemen.

Hoofdstuk 8 beschrijft de huidige praktijk van screening door jeugdverpleegkundigen, in het bijzonder de relatie tussen suïcidegedachten en gedrag en de beoordeling en vervolgacties van de jeugdverpleegkundigen. Voor dit doel zijn de gegevens van 3692 brugklassers geanalyseerd. De meerderheid van de uitnodigingen voor een preventief gezondheidsonderzoek werd niet gevolgd door een actie van de jeugdverpleegkundige. De jeugdverpleegkundige beoordeelde de meeste leerlingen met een aandachtspunt niet als risicovol. Een andere

bevinding is dat de zelfrapportage van suïcidegedachten een toegevoegde waarde heeft voor de voorspellende waarde van emotionele problemen zoals gemeten met de SDQ en andere indicatoren van gezondheid.

Ten slotte bevat **hoofdstuk 9** een algemene discussie van de bevindingen in dit proefschrift. Richtingen voor toekomstig onderzoek en implicaties voor de praktijk worden aangegeven. Een algemene conclusie is dat het effect van de verschillende methoden van dataverzameling beperkt is. Of we nu varieerden in de anonimiteit van de vragenlijst (anoniem of vertrouwelijk), de methode van dataverzameling (digitaal of schriftelijk), of het seizoen van dataverzameling, verschillen tussen groepen waren klein of niet significant. Over de vier studies die gericht waren op de validiteit van de meetinstrumenten is geen algemene conclusie te trekken.

Voor toekomstig onderzoek is het ondermeer aan te bevelen om studies uit te voeren naar de cross-culturele validiteit van de meetinstrumenten die gebruikt worden in jeugdgezondheidsonderzoek. Eveneens is het aan te bevelen om te onderzoeken of vragen voor leerlingen op het speciaal onderwijs aangepast moeten worden en te verkennen of andere methoden van dataverzameling beter geschikt zijn voor deze specifieke groep. Andere suggesties zijn om te onderzoeken welke factoren samenhangen met de zorgen van leerlingen over de vertrouwelijkheid en privacy, en de impact van het invullen van de Jeugdmonitor vragenlijst op het welbevinden van leerlingen.

Dankwoord

Op deze plek wil ik alle mensen bedanken die me tijdens het schrijven van dit proefschrift gesteund hebben, in welke vorm dan ook. Ik realiseer me terdege dat dit proefschrift er nooit had gelegen zonder al deze mensen!

Tijd om op deze plek ook even terug te blikken naar ruim vijf jaar geleden. Weer terug bij de GGD, na de geboorte van mijn dochter, informeerde mijn toenmalig leidinggevende, Erik Jan de Wilde, of ik belangstelling had voor een promotietraject binnen de toen nog op te richten academische werkplaats CEPHIR. Concreet betekende dit dat ik één dag per week de tijd hiervoor zou krijgen en een werkplek bij de afdeling Maatschappelijke Gezondheidszorg van het Erasmus MC. Toe aan meer wetenschappelijke verdieping leek me dit een mooie uitdaging en heb dus ja gezegd. De ideeën voor een proefschrift waren in het begin nog wel vaag. Ik zou me gaan verdiepen in het begrip 'welbevinden' en daarbij gebruik maken van de data die we bij de GGD met de Jeugdmonitor Rotterdam hadden verzameld. Al met al bleek het toch niet zo makkelijk om er een concreet onderwerp van te maken en gaandeweg ontstond het idee om te gaan focussen op methodologische aspecten van de Jeugdmonitor en de validiteit van de vragen die aan jongeren worden voorgelegd. Een zeer relevant onderwerp gezien de omvang aan gegevens die jaarlijks in het kader van de Jeugdmonitor Rotterdam verzameld worden bij jongeren. Als ik terugkijk op deze periode dan was het inderdaad een enorme uitdaging waar ik ontzettend veel van geleerd heb.

Op de eerste plaats bedank ik de GGD Rotterdam-Rijnmond voor de kans die mij geboden is om dit promotietraject in te gaan. Mijn beide promotoren, Prof.dr. Marianne Donker en Prof. dr. Frank Verhulst, bedankt voor de wijze waarop jullie mij begeleid hebben, jullie vertrouwen en jullie flexibele instelling op momenten waarop het proefschrift toch een iets andere wending kreeg. Erik Jan de Wilde, mijn co-promotor, bedankt voor je prettige begeleiding en het vertrouwen dat je mij hebt gegeven tijdens deze periode. Jouw positieve invloed en relativeringsvermogen droegen er vaak aan bij dat ik weer energie kreeg om verder te gaan. Ik zal de bezoekjes aan het Nederlands Jeugdinstituut gaan missen!

De leden van de kleine commissie, Prof.dr. Peter Muris, Prof.dr. Remi Hirasing en dr. Hein Raat, wil ik bedanken voor het lezen en beoordelen van mijn proefschrift. Prof.dr. Simone Buitendijk en Prof.dr. Edith de Leeuw, dank voor jullie toezegging om zitting te nemen in de grote commissie.

Wilma Jansen, collega èn nicht, bedankt voor je kritische commentaar op concepten en het delen van onze ervaringen tijdens dit promotietraject. Bij jouw promotie kon ik als paranimf al een beetje aan de sfeer proeven. Leuk dat jij nu mijn paranimf wilt zijn. Clothilde Bun, ex-collega, fijn dat jij mijn andere paranimf bent! Het is altijd heerlijk om met jou bij te kletsen over het reilen en zeilen bij de GGD'en waar we werken (en natuurlijk over allerlei andere zaken!).

Mijn collega's èn co-auteurs bij de GGD, Judith Goldschmeding, Cathelijne Mieloo en Christina Hoogeveen: dank voor jullie commentaar op conceptversies van de artikelen en de discussies die we daar over hebben gehad. Ook alle andere collega's bij de GGD en MGZ, teveel om allemaal persoonlijk te noemen, dank voor jullie belangstelling! In het bijzonder wil ik alle onderzoekers en onderzoeksassistenten van de sectie Jeugdmonitor en Onderzoek noemen. Frouwkje de Waart, sectieleider, dank voor je enthousiasme en steun. Alison Sutton, jij nog bedankt voor je correcties op mijn Engelse taalgebruik! Bij MGZ wil ik in het bijzonder Ineke Vogel noemen: Ineke, dank voor al je tips en adviezen en de gezelligheid op de vrijdagen!

Ik wil ook alle jeugdverpleegkundigen en doktersassistenten bedanken. Jullie hebben een belangrijk aandeel gehad in de dataverzameling, namelijk het afnemen van de vragenlijsten op scholen. Ook bedankt voor al jullie waardevolle opmerkingen! Verder wil ik alle scholen bedanken voor hun deelname en de jongeren voor het invullen van de vragenlijsten.

Waar ik ook met veel plezier op terug kijk is de samenwerking die we gezocht hebben met Prof.dr. P.D. Treffers en dr. A.W. Goedhart, de 'grondleggers' van de Nederlandse versie van de Strengths and Difficulties Questionnaire. Beste Flip en Arnold, ik herinner me nog goed ons eerste overleg bij het Curium in Oegstgeest. Het was in het begin zoeken wat we aan elkaar konden hebben, maar inmiddels ligt er een mooi artikel over de SDQ! Zonder jou, Arnold, had het artikel over de SDQ er niet gelegen. De manier waarop we aan het SDQ artikel hebben gewerkt, voornamelijk via de e-mail, vond ik bijzonder prettig. Er zijn enorm veel analyses uitgevoerd, zoveel dat ik soms door de bomen het bos niet meer zag. Arnold, veel dank voor deze prettige samenwerking. Ook Flip bedankt voor je reacties op het artikel vanuit Frankrijk!

Lieve vriendinnen en vrienden, zeker de afgelopen twee jaar waren de contacten wat mij betreft wat minder frequent. Ik moest vaak keuzes maken en dan koos ik er wel eens voor om eerder naar huis te gaan of toch maar niet dat ene nachtje nog te blijven slapen (dan kon ik namelijk de volgende dag nog iets doen….). Dank voor jullie belangstelling en luisterend oor! Wat de 'Ruisdaeltjes' betreft: met ons 'jubileum' in juni ben ik er echt het hele weekend bij!

Lieve ouders, jullie steun en hulp zijn niet in woorden uit te drukken! Jullie hebben ontzettend veel gedaan om er voor te zorgen dat ons gezinsleven op rolletjes bleef draaien. Het kwam ook regelmatig voor dat ik mijn gezin 'ontvluchtte' om bij jullie rustig te kunnen werken en me door jullie te laten verwennen met zelfgebakken appeltaart of appelbollen. De afgelopen vijf jaar stond ook in het teken van jouw ziekte, pa. Geen gemakkelijke periode, maar ik vind het heel bijzonder om te zien hoe positief jij in het leven staat. Ik bewonder jullie enorm en ik weet dat jullie erg trots op mij zijn.

Natuurlijk dank ik ook de rest van mijn familie: Marianne, Steven, Iris, Heleen, Andreas, Danielle, Alexander, Pieter, Anastasia, Svyatoslav, Annemarie en Peter, bedankt voor jullie

belangstelling en steun. Heleen, jij in het bijzonder bedankt voor al die uren die je gestoken hebt in het kritisch nalezen van het hele proefschrift. Ik had me geen betere 'editor' kunnen voorstellen! Daan en Alex, dank voor de leuke fotosessie en de lange reis die jullie hiervoor hebben moeten maken!

En tot slot, de mensen die me zeer dierbaar zijn: Joris, Tom, Max en Merel. Jullie zijn enorm trots op mij. Lieve Tom, zes jaar was je toen ik aan dit avontuur begon en nu ga je al bijna naar de middelbare school. Ik weet nog goed dat ik je vertelde over mijn eerste afspraak met een professor. Je was erg geïnteresseerd en vroeg of die professor echt een witte jas aan had en of hij grijze haren had en een brilletje op. Ik denk dat het beeld van een professor bij jou nog weinig veranderd is. Nou Tom, op 1 april kun je echte professoren ontmoeten en ik denk dat je je beeld dan toch moet bijstellen! Lieve Max, de uitspraak 'pluk de dag' past heel goed bij jou. Ik denk dat jij je er nooit druk over hebt gemaakt als ik op zondagochtend achter de computer zat. Het belangrijkste was dat jij dan aan het andere bureau, en vaak nog in je pyjama, computerspelletjes kon doen. Lieve vrolijke Merel, vijf jaar geleden nog een kleine baby en inmiddels heb je je ontwikkeld tot een prachtige kleuter. Jij hebt er allemaal maar weinig van meegekregen van een moeder die vaak nadat ze je naar bed had gebracht de studeerkamer invluchtte om nog wat te doen. Alhoewel, je had het laatst over 'proviveren, of hoe heet dat wat jij doet?', dus het gaat blijkbaar niet helemaal langs je heen.
Allerliefste Joris, de laatste persoon om te bedanken ben jij! Jij weet als geen ander hoe ik deze periode heb meegemaakt. Het was een superhectische periode, zeker als je zo in elkaar zit dat je op alle fronten aan ieders verwachtingen wilt blijven voldoen (wat natuurlijk onmogelijk is). Een dag per week aan een proefschrift werken is niet veel, en dat betekende dan ook dat veel weekenddagen en vakantiedagen aan het proefschrift besteed zijn, zeker de afgelopen twee jaar. Dank voor al je steun en je opbeurende adviezen, vooral als het eens tegen zat. Jouw optimistische kijk op het leven hebben er aan bijgedragen dat ik deze eindstreep heb gehaald. De laatste stelling is voor jou!

List of publications

This thesis

Van de Looij-Jansen PM, Goldschmeding JEJ, de Wilde EJ (2006), Comparison of anonymous versus confidential survey procedures: effects on health indicators in Dutch adolescents. *Journal of Youth and Adolescence* 35: 659-65.

Van de Looij-Jansen PM, de Wilde EJ (2008), Comparison of web-based versus paper-and-pencil self-administered questionnaire: effects on health indicators in Dutch adolescents. *Health Services Research*: 1708-21.

Van de Looij-Jansen PM, de Wilde EJ, Mieloo C, Donker MCH, Verhulst FC (2009), Seasonal variation in self-reported health and health-related behavior in Dutch adolescents. *Public Health,* doi:10.1016/j.puhe.2009.07.015.

Jansen W, **Van de Looij-Jansen PM**, Ferreira I, de Wilde EJ, Brug J (2006), Differences in measured and self-reported height and weight in Dutch adolescents. *Annals of Nutrition and Metabolism* 50: 339-346.

Van de Looij-Jansen PM, Goedhart AW, de Wilde EJ, Treffers PDA Confirmatory analysis of the factor structure and factorial invariance of the self-report SDQ: how important are method effects and minor factors? *British Journal of Clinical Psychology (in revision).*

Van de Looij-Jansen PM, Jansen W, de Wilde EJ, Donker MCH, Verhulst FC. Discrepancies between parent-child reports of internalizing problems among preadolescent children: relationships with gender, ethnic background and future internalizing problems. *Accepted for publication by Journal of Early Adolescence.*

de Wilde EJ, **Van de Looij-Jansen PM**, Goldschmeding JEJ, Hoogeveen WC. Self-report of suicidal thoughts and behaviour and school nurse evaluations in Dutch high school students. *Submitted*

Other publications

Van de Looij-Jansen PM, Bun CJE, Butte D, de Wilde EJ (2003), De samenhang tussen psychisch welbevinden en gezinsfactoren bij Turkse en Nederlandse adolescenten. *Tijdschrift voor Gezondheidswetenschappen*: 189-195.

Ezendam NPM, Oenema A, **Van de Looij-Jansen PM**, Brug J (2007), Design and evaluation protocol of "FATaintPHAT", a computer-tailored intervention to prevent excessive weight gain in adolescents. *BMC Public Health*, 7: 324.

Jansen W, **Van de Looij-Jansen PM**, de Wilde EJ, Brug J (2008), Feeling fat rather than being fat may be associated with psychological well-being in young Dutch adolescents. *Journal of Adolescent Health* 42: 128-36.

Van der Horst K, Oenema A, **Van de Looij-Jansen P.M.**, Brug J (2008), The ENDORSE study: research into environmental determinants of obesity related behaviors in Rotterdam schoolchildren. *BMC Public Health* 8: 142.

Junger-Tas J, Cruyff MJLF, **Van de Looij-Jansen PM**, Reelick NF (2003), *Etnische minderheden en het belang van binding. Een onderzoek naar anti-sociaal gedrag van jongeren.* Den Haag: SDU Uitgevers B.V.

Van Bergen DD, Eikelenboom M, Smit JH, **Van de Looij-Jansen PM**, Saharso S Suicidal behavior and ethnicity of young females in Rotterdam, the Netherlands: rates and risk factors. *Ethnicity and Health (submitted)*

Van den Einde-Bus AEM, Goldschmeding JEJ, Tielen LM, de Waart FG, **Van de Looij-Jansen PM** (2010), De gezondheid van jongeren met een langdurig ziek, gehandicapt of verslaafd familielid: reden tot zorg. Accepted for publication in *Tijdschrift voor Gezondheidswetenschappen*.

Over de auteur

Petra van de Looij-Jansen werd op 2 februari 1967 geboren in Arkel. Ze behaalde in 1985 haar VWO diploma aan het Gymnasium Camphusianum te Gorinchem. In datzelfde jaar startte ze in Nijmegen met de opleiding Voeding en Diëtetiek. In 1990 begon ze de opleiding Voeding aan de Landbouw Universiteit te Wageningen en koos hierbij als specialisatie Voeding en Gezondheid. Haar afstudeeronderzoeken verrichtte ze bij het Nederlands Kankerinstituut in Amsterdam waar ze onderzoek deed naar de relatie tussen voeding en het risico op borstkanker, en bij het St. John's Medical College in Bangalore (India).

Na haar afstuderen begon ze als onderzoeker / epidemioloog bij het cluster Jeugd van de GGD Rotterdam-Rijnmond. Ze voerde diverse onderzoeken uit op het terrein van de jeugdgezondheidszorg. In 1996 stond ze aan de wieg van de Jeugdmonitor Rotterdam, een instrument dat in de Rotterdamse regio ingezet wordt om lichamelijke, emotionele en gedragsproblemen bij kinderen en jongeren vroegtijdig te signaleren, zowel op individueel als op collectief niveau. In de afgelopen jaren heeft ze bijgedragen aan de verdere ontwikkeling, onderbouwing en implementatie van dit unieke instrument.

Eind 2004 startte ze in deeltijd haar promotieonderzoek met als onderwerp het gebruik van zelfrapportage vragenlijsten in jeugdgezondheidsonderzoek. De onderzoeksvragen voor haar proefschrift vloeiden voort uit de dagelijkse praktijk van de Jeugdmonitor Rotterdam. Ze voerde het promotieonderzoek uit in het kader van CEPHIR (Centre for Effective Public Health In the Rotterdam area), de academische werkplaats Publieke Gezondheid in de regio Rotterdam-Rijnmond. De resultaten staan beschreven in dit proefschrift.

PhD Portfolio Summary

Summary of PhD training and teaching activities

Name PhD student: Petra van de Looij
Erasmus MC Department: MGZ
Research School:

PhD period: 2004-2009
Promotor(s): Prof.dr. M.C.H. Donker en Prof.dr. F.C. Verhulst
Supervisor: Dr. E.J. de Wilde

1. PhD training	Year	Workload (Hours/ECTS)
General academic skills		
- How to write a medical paper, Erasmus Summer Programme	2004	30
Research skills		
- Multilevel-analyse (SPSS)	2006	8
Presentations		
- Symposium kinder- en jeugdpsychiatrie, Boerhave Commissie Leiden (2006) Gebruik van de SDQ in de Jeugdmonitor Rotterdam. De introductie van de Strengths and Difficulties Questionnaire (SDQ) en daarmee gerelateerde instrumenten in Nederland.	2006	20 20
- Nederlands Congres Volksgezondheid (2007), Emotionele problemen bij 9-10 jarige kinderen: voorspeller van emotionele problemen drie jaar later?	2007	20
- Posterpresentatie Nederlands Congres Volksgezondheid (2007), Een web-based of papieren vragenlijst bij jongeren: verschil in beantwoording?	2007	20
- Posterpresentatie Nederlands Congres Volksgezondheid (2008), Speelt seizoensvariatie een rol bij jeugdgezondheidsenquêtes?	2008	20
- Nederlands Congres Volksgezondheid (2009), Gezondheid van ROC deelnemers: reden tot zorg?	2009	20
Seminars and workshops		
- Benelux-Nordic seminar. Indicators of Child and Youth Well-being. The link between knowledge, policy and practice.	2009	10
- Jeugd in Onderzoek. De kracht van preventie in de jeugdketen	2009	8
- Puberproblemen. De normaalste zaak van de wereld? Onderkenning en signalering in het voortgezet onderwijs	2007	8
Other		
- Van Onderzoek naar Beleid, Netherlands School of Public and Occupational Health (NSPOH)	2008-2009	80
- Projectmatig creëren (Pontus Bridging)	2007	20

Erasmus MC
Universitair Medisch Centrum Rotterdam

2. Teaching activities	Year	Workload (Hours/ECTS)
Supervising Master's theses		
- Seasonal variation in youth health surveys (Internship Epidemiology and Public Health, Wageningen University)	2007	30
Other		
- Training/instructiebijeenkomsten over uitvoering Jeugdmonitor Rotterdam aan jeugdverpleegkundigen GGD Rotterdam-Rijnmond	2004-2009	50
- NIHES studenten: presentatie JMR	2008	5