



Subjective health complaints in adolescence— Reliability and validity of survey methods

SIREN HAUGLAND AND BENTE WOLD

This paper studies test–retest reliability and validity of one measure of adolescent health complaints. The test–retest included an eight-item symptom checklist developed for the survey of *Health Behaviour in School-aged Children* ($n=344$). Qualitative analysis showed adequate validity for most items. For the total sample, all items were found to have adequate intraclass correlation coefficients (ICC) in the range 0.61–0.75. There were inter-item differences and girls generally received the higher values. Most changes were within one category. Adolescents' understanding of 16 complaints was studied by interviews with 38 adolescents. A few items showed ambiguity in interviews despite adequate test–retest stability.

© 2001 The Association for Professionals in Services for Adolescents

Introduction

Adolescent health is most often studied by measures of subjective or perceived health. Scales and symptom checklists are commonly used measurement formats, as are single item measures where young people are asked to rate their health as “excellent, good, fair or poor”. These measures have largely captured domains defined by adults and practitioners, and given knowledge of symptom prevalence rather than qualitative aspects such as impact on functional ability and well being. The use of symptom checklists in adolescence is based on three underlying assumptions. First, adolescents are at a level of cognitive development where they are able to differentiate concepts of health and illness. Second, this age group is able to understand the content of symptom checklists, and third, adolescents can evaluate and report feelings and complaints reliably. Previous studies found that concepts of health and illness are differentiated throughout childhood and adolescence (Millstein, 1993) and suggest that developmental factors may influence both reliability and validity of subjective health measures. Based on interviews and test–retest data, the purpose of the present paper is to study test–retest reliability and validity of one measure of subjective health complaints in two samples of 14–16 year olds.

Subjective health complaints

Subjective health complaints refer to symptoms experienced by the individual with or without a defined diagnosis. Such symptoms constitute both everyday experiences and health problems, and are common causes of disability and sickness certificates in adults (Tellnes *et al.*, 1989). For many symptoms the subjective report is the only measure that both clinicians and researchers rely upon, as these are largely “unmeasurable” in a clinical sense.

Previous studies show that a large number of adolescents report weekly health complaints, and that such symptoms frequently are the cause of absence from school (Mikkelsen *et al.*, 1997). The reporting of symptoms is prevalent already at the age of eleven years (Haugland,

Reprint requests and correspondence should be addressed to: Siren Haugland, Research Center for Health Promotion, University of Bergen, Christiesgt 135015 Bergen, Norway. (E-mail: Siren.haugland@psych.uib.no).

2001). There is clustering of symptoms, with one group of adolescents reporting several symptoms (Pennebaker, 1982; Starfield *et al.*, 1984; Alfvén, 1993; Klepp *et al.*, 1996; Mikkelsson *et al.*, 1997). Headache, abdominal pain, irritability, and nervousness are the symptoms most often reported (King *et al.*, 1996). In adolescence, most studies find that girls report more symptoms than boys do and both prevalence of complaints and gender differences increase with age (Aro, 1987; King *et al.*, 1996; Klepp *et al.*, 1996; Haugland *et al.*, 2001). Cross-national studies demonstrated that the level of reported health complaints varied between countries, but the patterns according to age and gender were consistent (King *et al.*, 1996). This study from 24 countries found that 23–53 per cent of 15-year-old girls and 14–30 per cent of 15-year-old boys reported having headaches weekly or more often, in the past six months. Similar figures were found for other health complaints and in other studies (Goodman and McGrath, 1991; Kolip and Hurrelmann, 1995).

Measuring subjective health complaints

Social, dispositional, and contextual variables were previously found to affect how symptoms and sensations were interpreted, whether these were reported to others, and the language used to describe them (Cioffi, 1996). Previous studies in adults suggest that there may be gender differences in reporting health complaints due to differences in the processing of visceral and somatic sensations (Pennebaker and Roberts, 1992). Men's visceral and somatic sensations were found to be more consistent over time and more independent of their current circumstances or situations. According to this, men's reporting of health complaints may be more stable than that of women. Women were found to use more situational cues and more external information than men to interpret bodily sensations. This could lower their threshold for bringing such complaints to the attention of others or reporting them in surveys (Wool and Barsky, 1994; van Wijk and Kolk, 1997).

Previous studies found that consistency in the reporting of subjective health was related to age. In a study of psychiatric self-reports by 6–12-year-olds, stability was found to be linearly dependent on age (Fallon and Schwabb-Stone, 1994). This relationship between age and consistency was also demonstrated in a sample of 9- and 14-year-olds (Otter *et al.*, 1995). Studies carried out using different symptom checklists and age groups suggest that adolescents are capable of reporting symptoms and complaints reliably. A test–retest study of musculoskeletal pain in a sample of 9–12-year-olds demonstrated a kappa of 0.9 with an interval of about one week (Mikkelsson *et al.*, 1996). Wisniewski *et al.* (1988) carried out a test–retest on Children's Psychosomatic Symptom Checklist in a sample of 11–14-year-olds. They found that item–total correlation was generally significant and exceeded 0.60. There were no significant gender differences or differences between the 1- and 5-week retests. Pennebaker (1982) carried out a test–retest of the PILL (Pennebaker Inventory of Limbic Languidness) in a sample of college students and found a reliability coefficient of 0.83 over a 2-month period. The item reliability with a 2-week interval averaged across all 54 items was 0.73 (Pennebaker *et al.*, 1977), demonstrating reliability in the studied age group.

Most often, testing for validity is seen as demonstrating the psychometric properties of a scale, and the topic is divided into content validity, criterion validity, and construct validity. Led by Cronbach (1971), the focus of validity testing changed to include also the characteristics of the people who were assessed. Thus, validation processes aim to determine the degree of confidence that can be placed on the inferences made about people based on their scores on a scale (Streiner and Norman, 1995). Symptom checklists are thought to be easy to understand, yielding satisfactory face validity. Few have studied other aspects of

validity in adolescence; thus, little is known about the perception of health complaints reported in surveys. Mikkelsen *et al.* (1997) found adequate observed agreement when comparing the results of a symptom checklist of non-specific musculoskeletal pain with those of interviews. However, Starfield *et al.* (1993) demonstrated inconsistencies in the understanding of one symptom checklist when focus group interviews were studied to assess content and construct validity of the Child Health and Illness Profile. The findings led to changes in wording, time frame, and response categories of most items adopted from other instruments. Testing for validity also resulted in new items concerning energy level and the feeling of “being on top of things”, as well as physical signs of good health (e.g. alertness, sparkling eyes, and clear skin). These findings demonstrate the need to study age-specific validity.

The HBSC-symptom checklist (HBSC-SCL)

The HBSC-symptom checklist is an eight-item symptom checklist developed for the survey of *Health behaviour among school-aged children (HBSC)* (King *et al.*, 1996). The study was established in 1983 (Aaroe *et al.*, 1986) and is currently carried out in close to 30 countries. This is a WHO cross-national survey where the aim is to obtain an understanding of young people’s health behaviour, lifestyle and perception of health. Thus, the aim is to make group inferences in studies of health complaints and their correlates, rather than employing the scale for diagnostic purposes. Table 1 illustrates the items of the HBSC-symptom checklist (HBSC-SCL). The table also lists eight additional items added in the qualitative study to explore the feasibility of including more items in the scale. The adolescents were asked how often they experienced these symptoms during the previous 6 months; about every day, more than once a week, about every week, about every month and rarely or never. The health complaints studied were thought to be well defined and to measure experienced symptoms. Most of the items listed were previously used in other measures of adult and adolescent health status (Pennebaker, 1982; Wisniewski *et al.*, 1988; Goodman and McGrath, 1991; Garralda, 1992; Klepp *et al.*, 1996).

Against this background, and focussing on gender differences, the purpose of the present paper is to study content validity and test–retest reliability of both an eight-item and extended version of the HBSC-symptom checklist, to answer the following questions: (1) What is measured by the symptom checklist? (2) Do these symptoms cover the phenomenon of health complaints in 14–16 year-olds? and, (3) How reliable are the results of survey methods in this age group?

Table 1 *HBSC-symptom checklist and additional items*

HBSC-SCL	Additional items
Headache	Neckache
Abdominal pain	Afraid
Backache	Exhausted
Dizziness	Heart beating fast
Feeling low	Nausea
Irritable/bad temper	Vomiting
Nervous	Limb pain
Sleeping difficulties	Bodyache

Method

Samples

Both samples were drawn from Norwegian schools that participated in the European Network of Health Promoting Schools (ENHPS) project from 1994 to 1997 (Burgher *et al.*, 1999). The project had a broad approach which attempted to address knowledge and skills development, as well as the physical and social environments. As part of the ENHPS project, the pupils took part in a survey twice a year. The latest survey was conducted 5 months prior to the present studies. The Norwegian ENHPS questionnaire covered questions of health behaviour, subjective health, and the social environment of the school. It also included the scale of subjective health complaints included in the present study. Apart from this questionnaire, no particular emphasis was placed on the concept of health or health complaints.

Sample 1. The qualitative sample consisted of ninth graders from one school outside Bergen, Norway. The community has a mixed urban/rural population. The pupils were contacted by a letter addressed to them and their parents, which was distributed by the teacher. They were informed that the purpose of the study was to gain insight into the perceptions of health and common complaints in adolescence. It was pointed out that each pupil's personal view was the matter of interest. They were also informed that the interviewer would be a medical doctor whose main task is research. The interviewer was not known to any of the pupils. The total sample consisted of 48 pupils, from two separate classes (class A, 22; class B, 26 pupils). Written consent was granted from the parents of 38 pupils. Two pupils did not get consent from parents and eight pupils did not reply (four pupils in each class). The final sample consisted of 38 ninth graders. Of these 20 were girls and 18 were boys. Thirty pupils were 15 years old and eight pupils were 16 years old. The 10 pupils missing were not thought to be different from the respondents (information from the teacher after the interviews were carried out).

Sample 2. A test-retest was conducted on a convenience sample of 344 adolescents aged 14–16 years old. This included 48 per cent (164) boys and 52 per cent (180) girls. The sample came from four different schools. Written consent was obtained from the parents of all participants.

Interview guide

Responses given to the following questions will be described in the present paper: (1) Do you experience any of these health complaints? (2) Do you experience any other complaints? If health complaints were experienced, they were asked to comment on "how often" and "what is it like for you to experience this health complaint".

Data collection

The qualitative data was collected in April and May 1997. The pupils were interviewed individually by the first author of this paper and each interview averaged 20 minutes. The interviews were carried out at the school during normal school hours. Previous information to pupils and parents was repeated before the interviews started (purpose of study, matter of interest). They were assured that all information given to the interviewer would remain confidential. They were informed that: (1) The main part of the interviews would be standardized and open-ended, but with an aim to build these as a dialogue where the focus

was the informant's understanding of the subject areas. (2) Any information they found relevant was of interest to the interviewer and that there were no right or wrong answers. (3) The interviewer would not express disagreement or other views on the information given. (4) There was no time limit and room would be left for informal conversation.

The informants were shown a list of health complaints containing 16 items, illustrated in Table 1. The items were shown in the same order for all interviews, deviating from this order only when the informants spontaneously explained about other symptoms. This ensured systematic information on all items of the symptom checklist. The informal conversation was included to be able to pursue emerging information, e.g. about additional complaints. The fact that data was collected by one interviewer may limit bias due to varying sensitivity to the themes investigated, and facilitate the same level and aspects of information from all informants, thus ensuring reliability in the traditional sense. However, from a phenomenological-hermeneutical standpoint it may also be seen as an advantage to have more than one interviewer, as this may give a broader picture of the themes focused upon (Patton, 1990).

The test-retest and interviews were conducted during normal school-hours at the school. The test-retest was carried out in May 1998, with a 1-week interval between test and retest. Data was collected through self-completion questionnaires, which were precoded. The teacher administered the surveys in the classroom, and the pupils returned the questionnaires in sealed envelopes. The test and retest were identical procedures, and the students did not know about the forthcoming retest when the test was administered.

Analysis

Content validity

Analysis of the qualitative data was conducted by the first author, and according to the principles described by Patton (1990). Each interview was audiotaped and transcribed in full by the interviewer. The analysis can be described as a three-phase process: description, reduction, and interpretation. The interviews were read several times to understand the content, and to take mental note of subject areas and context. Responses were recorded for each interview to identify themes and primary patterns in data. "Recurring regularities" represent patterns that can be sorted into categories (Patton, 1990). The responses were coded according to each health complaint and the themes mentioned by the informants and focusing on the subject's interpretations. Case analyses were studied first, where the interview of each informant was studied separately using all of the data for each person. Variations in individuals were studied on this basis. A cross-case analysis was then carried out grouping together answers from different informants to each question and health complaint. This part of the analysis led to a description of variations in answers to the list of health complaints. The third phase of the analysis was to interpret the meaning of the emerging categories based on recurring features of the phenomenon under study. The pupils were interviewed in Norwegian and the responses cited here were translated into English after the analyses.

Test-retest reliability

Pearson's r is often used as an estimate of reliability. However, it has been argued that reliability is inflated by the use of this measure (Streiner and Norman, 1995). Several authors

have advocated the use of intraclass correlation (ICC) as a measure of test–retest stability, as this considers the variance between test and retest (Shrout and Fleiss, 1979). In this study, ICC was estimated by the use of a two-way mixed consistency model, and computed for the eight item sum-score, as well as single items of the HBSC-SCL.

Analysis was also repeated for an extended eleven-item version of the HBSC-SCL (Table 1). All analyses were repeated for the total sample and for boys and girls separately. The 95 per cent confidence interval was computed for the differences between genders (Table 2). Pearson's r was calculated to be comparable with ICC, but for the total sample only. The number of pupils who showed no difference between test and retest determined the stability rate. The items studied are at the ordinal level of measurement. Thus, it is possible to grade both perfect and partial stability. A deviation of one category between test and retest is considered more stable than that of two or more categories. Stability for each item was determined by the subtraction of (testscore–retestscore). The combined results for perfect and partial stability provide an overall picture of the stability for each item. Chi-square tests were carried out to study the association between stability and gender. Statistical analyses were conducted by means of SPSS for Windows (version 9.0).

Results

Content validity

The responses given reveal that most of the items on the 16-item list of health complaints (Table 1) seem to qualify for the term “complaint”. The results show that the HBSC-symptom checklist is likely to reflect differences between children who experience health and well being and those who do not. The adolescents' comments suggest that the health complaints studied influence functional ability. The informants were asked if they experienced any other complaints than the ones listed, and the following complaints were mentioned: hay fever, other allergies, asthma, weight problems, ear problems and shoulder pain. Other symptoms were perceived as accurately described in the symptom checklist.

Except for period pain there were no gender differences in the informants' experience of the health complaints listed. However, a few items were not perceived as complaints. “Heart beating fast” was related to normal activities and positive experiences in the lives of adolescents, e.g. being in love, “*If I see a very handsome boy I get that*”. “Dizziness” seems to constitute a category largely experienced as normal physiological change, e.g. dizziness after getting up quickly. “Nausea” and “vomiting” were commented on by few pupils and were most often related to dietary factors, either hunger or too much food. “Sleeping difficulties” were seen as a result of lifestyle, e.g. sleeping during the day.

Previous factor analysis of the eight item HBSC-SCL favoured a model of two correlated factors where one factor can be labelled somatic and one represents psychological symptoms (Haugland *et al.*, 1999). This distinction corresponds with the understanding of the informants in this study, and further symptom description will be presented according to the two factors. Additional items are included in keeping with the comments given by the informants.

Somatic complaints

Consistent with previous quantitative studies, “headaches”, “backache”, and “neckache” were frequently reported (Goodman and McGrath, 1991; King *et al.*, 1996; Mikkelsen, 1997), and often led to the use of medical services. Perceived aetiology was often commented on and seen as related to lack of activity (e.g. sitting at school), type of activity (e.g. using a computer), workload (e.g. at school), and defined diagnoses (e.g. migraine, scoliosis, Mb Scheurman, whiplash).

“I have quite a lot of headaches, ever since I started secondary school, with stress and things like that.”

“I have headaches once in a while. It’s probably because of the heavy air in the classroom.”

“Because it (backache) is either growing pains or from sitting in a strange position.”

“Backache several times a week. It’s probably because my school bag is too heavy.”

“I do have neckaches. If I sit a lot and write I get that. It’s especially at weekends, after the week. . . I can really feel it. Stiff all over.”

Few pupils reported “abdominal pain”. This symptom was most often described as period pain or pain after eating. However, abdominal pain was also mentioned in relation to negative experiences with family, school or peers:

“You never know what it’ll be like to come home, so then you have constant stomachaches.”

“Limb pain” and “bodyache” seem to be overlapping constructs, where few pupils experienced these as general states. Most of the informants experienced such symptoms after exercise. “I get limb pain after getting hurt.” “I have sore muscles if I haven’t stretched properly (after exercise).”

A few informants described growing pains. Limb pain and bodyache was also described as a general feeling or state:

“Sometimes you feel your body is totally dead.” “It (bodyache) is not really painful, but I feel sometimes I’m really weak and miserable.”

Psychological complaints

“Feeling low” was reported to be an emotional state lasting for hours, days and weeks, caused by known or unknown factors. This feeling influenced daily activities and the informants’ relationships with other people.

“It very much depends on how school is. Sometimes I feel no one wants to talk to me and that no one likes me and stuff like that. And then I feel low. When I have these periods it (feeling low) often lasts for more than a day or two.” “Most often that feeling lasts for two weeks or so. N has told me that there are many others both in my class and otherwise who feel the same way as I do. And then I feel a little better, when I don’t feel like I’m all alone in the world.”

“I don’t feel as low every day, but it’s there almost every day.”

“Sometimes I’m a bit sad. I don’t think there’s any reason, but I just feel it. It’s a bit more like that when I get up. . . . crappy day, I think. And then the next day it’s the same. . .”

“Irritability and bad temper” were complaints commonly reported. These were described as transitory feelings or experiences, rather than a lasting state. These symptoms were not reported as influencing life and daily activities in the same way as “feeling low”. Symptoms

were seen as related to others (e.g. family, friends) or themselves (e.g. puberty, tired). *"Grumpy, I feel it myself actually. Want to protest about everything and want everything my own way."*

"Afraid" and "nervous" were largely seen as interchangeable, of which nervous seems to be the "preferred" item. Nervousness was experienced as a reaction to a situation or stressor, rather than a state that lasted over time. There were two sub-categories, of which one represents a "normal" fear of the unknown; *"I'm a bit nervous when I go somewhere and meet new people"*. Nervousness was also experienced as a result of known activities or expectations from parents, teachers and friends. This type of nervousness was reported in situations when they had to stand in front of others, such as reading in the classroom.

"... if I'm to do something new ... then I think of all sorts of strange things, maybe I'll do something wrong, and then people will laugh, or something like that." "When I'm standing in front of the classroom and talking to the others."

"Afraid" was not a common complaint and was largely experienced as being afraid of the dark. No students described being afraid as a general feeling state, which lasted over time. *"I'm very afraid of the dark. So I don't go out alone in the dark."*

However, one group described a feeling that occurred when the informant had to stand in front of others. *"I am afraid when we work together with other teams that I don't know. It's a bit embarrassing when I have to go out front and show everyone what we have to do."*

The comments on "exhausted" can also be divided into two subcategories. The first category was experienced as a transitory state and a feeling that influenced everyday life in a negative way. School was listed as a main cause of this state.

"And school makes you feel exhausted. At the moment I don't manage to do anything when I get home. And then it gets better during the evening." Exhaustion was also experienced as a more general state of fatigue related to other negative feelings and even self-destructive thoughts. The informant in the following statement felt exhausted quite often and, when asked what it was like, answered: *"I'm sick of school and life and everything. I have thought about that sometimes. To end my life."*

Test-retest reliability

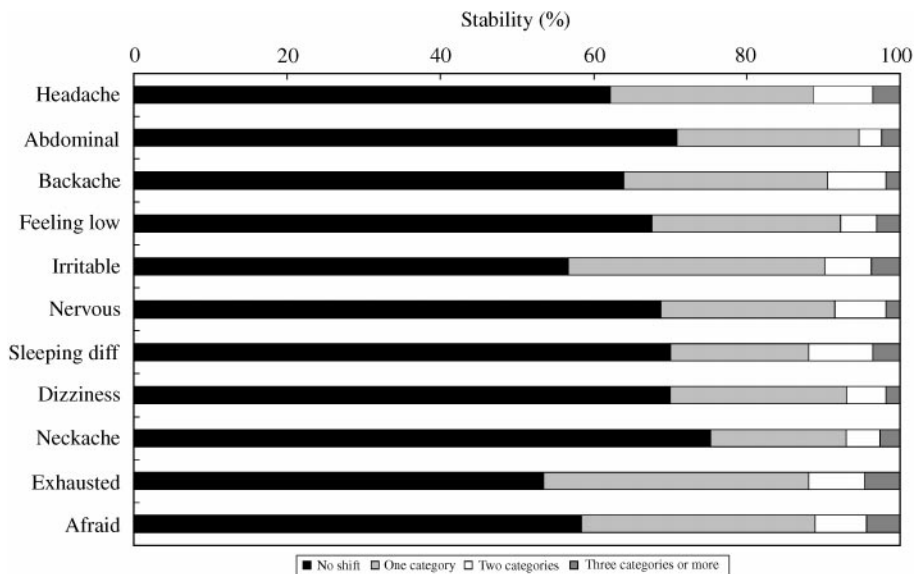
Intraclass correlation coefficients (ICC) were computed for each item separately and for the sum-scores of the eight-item symptom checklist. By convention, reliability coefficients of 0.80 would be considered as good test-retest reliability. Lower coefficients are considered adequate on data designed for group level inferences, such as the present study. All analyses were carried out for the total sample and for boys and girls separately. Table 2 presents stability based on intraclass coefficients. The 95 per cent confidence intervals are listed for boys and girls separately and for gender differences. Pearson's r is shown for the total sample. The sum-score of the eight-item HBSC-SCL was found to have an ICC of 0.79 for the total sample, 0.76 for boys, and 0.81 for girls. The corresponding values for an extended eleven-item version were 0.75, 0.77 and 0.83, respectively.

The stability of single items on the symptom checklist varied to a large degree. For the total sample, all items were found to have adequate ICC in the range 0.61–0.75. Girls generally received higher values than boys both for the total score and for most single items. The exceptions were "nervous" and "neckache" where boys had the higher values. The 95 per cent confidence interval for gender differences demonstrated significance at the 5 per

Table 2 Test–retest stability of total sample/according to gender

Scale/Item	Boys (n=164)		Girls (n=180)		Gender difference 95 per cent CI	Total (n=344)	
	ICC	95 per cent CI	ICC	95 per cent CI		ICC	Pearson's r
HBSC-SCL (8)	0.76	(0.68–0.81)	0.81	(0.75–0.86)	(–3.6–13.6)	0.79	0.79
HBSC-SCL (11)	0.77	(0.70–0.83)	0.83	(0.78–0.87)	(–2.4–14.4)	0.81	0.81
Headache	0.62	(0.51–0.70)	0.69	(0.61–0.76)	(–3.0–17.0)	0.67	0.67
Abdominal pain	0.65	(0.55–0.73)	0.68	(0.59–0.75)	(–4.0–13.0)	0.67	0.67
Backache	0.69	(0.60–0.76)	0.76	(0.69–0.81)	(–2.4–16.4)	0.73	0.73
Feeling low	0.51	(0.39–0.61)	0.74	(0.67–0.80)	(13.0–33.0)*	0.67	0.67
Irritable	0.57	(0.46–0.66)	0.66	(0.57–0.74)	(–1.3–19.3)	0.61	0.61
Nervous	0.70	(0.61–0.77)	0.64	(0.54–0.72)	(–4.0–16.0)	0.67	0.67
Sleeping diff	0.71	(0.63–0.78)	0.77	(0.70–0.82)	(–3.3–15.3)	0.75	0.76
Dizziness	0.67	(0.58–0.75)	0.78	(0.71–0.83)	(1.6–20.4)*	0.73	0.74
Neckache	0.80	(0.74–0.85)	0.73	(0.66–0.79)	(–1.9–15.9)	0.76	0.76
Exhausted	0.65	(0.55–0.73)	0.68	(0.60–0.75)	(–7.0–13)	0.67	0.67
Afraid	0.63	(0.53–0.71)	0.73	(0.65–0.79)	(0.2–19.8)*	0.68	0.68

*Significant at the 5 per cent level.

**Figure 1.** Item stability for the total sample (%).

cent level for “feeling low”, “dizziness”, and “afraid”. Differences of other items were approaching significance. For girls, all single items showed acceptable levels of reliability within the range 0.64–0.78. For boys, most items were found to have adequate stability. The items of “feeling low” and “irritable/bad temper” had ICC below 0.60, which is lower than desired for psychometric purposes. Both boys and girls were found to have a high ICC for the item “neckache”. Intraclass correlation and Pearson’s *R* were interchangeable.

Figure 1 shows item stability for the total sample. The number of adolescents who showed perfect stability varied between 54 and 75 per cent, for “exhausted” and “neckache”, respectively. Most stability rates were in the range 65–70 per cent and most changes were within one category. For the item “neckache”, boys showed a significantly lower number of category changes than girls did (Chi-square = 14.4, df (4), $p=0.006$). For other items there were no significant gender differences.

Discussion

Content validity

The present study found encouraging evidence for the validity of subjective health measures used in the studies of *Health Behaviour in School aged Children* (HBSC) and *The European Network of Health Promoting Schools* (ENHPS). All items were found to have good face validity. The qualitative data supports the findings of quantitative studies, where a large number of adolescents reported on subjective health complaints (Pennebaker, 1982; Belmaker, 1984; Aro and Taipale, 1987; Goodman and McGrath, 1991; Kolip and Hurrelmann, 1995; King *et al.*, 1996; Klepp *et al.*, 1996). The comments made by the informants suggest that health complaints have a negative influence on subjective well-being and functional ability in the age group studied. Gender differences in validity seem to be limited to the level of symptoms reported rather than the understanding of symptoms or impact on everyday life. This consistency suggests that boys and girls may have learned a common interpretation, and supports studies where pain is seen as a reflection of cultural norms (Wool and Barsky, 1994; Bendelow and Williams, 1995; vanWijk and Kolk, 1997). In the sample studied, the interview approach may be seen as more “gender neutral” than the questionnaire.

Current thinking on somatic interpretation is often biased toward viewing awareness of symptoms as necessarily distressing (Cioffi, 1991). The items included in the present study seem to represent different constructs both in nature and influence on daily activities. For the item “heart beating fast” the term health complaint does not seem to be applicable, as this was associated with the more positive feeling of being in love. This illustrates the need to validate symptom checklists in the target population. The items “afraid” and “exhausted” showed adequate stability, but the interviews revealed some ambiguity when describing these items. Musculoskeletal symptoms have been a focus of interest in different age groups. In the present study, the items of “backache” and “neckache” constitute musculoskeletal health complaints as these are most often defined, and suggest that “neckache” should be added to the core items of the HBSC-SCL. The items of “limb- and bodyache” seem to constitute a category of complaints that largely occur after exercise (sore muscles) and are not perceived as causes of distress. This is in keeping with previous studies (Kujala *et al.*, 1999), and studies where somatic sensations due to physical activity were examined. When asked to attend to such sensations during exercise these were predominantly perceived as neutral or pleasant (Cioffi, 1991).

Reliability

This study showed adequate test-retest reliability for the sum-scores and most single items of the symptom checklist used in the study of *Health Behaviour in School aged Children* (HBSC-SCL). However, there were striking inter-item differences both in the pattern of category

change and intraclass coefficients (ICC). Adolescents in this sample were asked to report frequency of symptoms in the past 6 months. Some variance between test and retest was to be expected, as experiencing the symptoms on the day of either test or retest may have influenced the category reported. Most changes were found to be within one category, indicating stability of the phenomenon under study and adequate scaling properties. In terms of reliability, the present study demonstrated adequate coefficients, with minor differences between the eight- and eleven-item versions. The test–retest reliability showed lower coefficients than previously demonstrated on the PILL by Pennebaker (1982). However, Pennebaker's study included 54 items and was carried out in an older sample of adolescents. Thus, reliability coefficients are not directly comparable as these are a function of test length (Ferguson, 1989).

Inter-item differences in reliability may be influenced both by aetiology and expression of symptoms. The perception of symptoms that are vague, mild, and ambiguous tend to be more variable and subjective than the perception of symptoms that are severe or due to obvious external causes (Mechanic, 1983). For other health complaints there is more room for psychological factors such as past experiences, mood, and other people's opinion to influence the reporting of symptoms. Thus, symptoms that refer to well-defined areas of the body were thought to be the more stable items. The present study only partially supports this hypothesis. A defined item such as "neckache" yielded high ICC and a high rate of adolescents showing absolute agreement. However, the results for less defined items were more diverse. "Irritability", "exhausted" and "afraid" showed the lowest rates of absolute agreement and the greatest tendency of a change of three or more categories, whereas symptoms such as "sleeping difficulties" and "dizziness" yielded a relatively high rate of absolute agreement and ICC in the upper range.

In the present study, girls generally yielded higher ICC than boys. Gender differences were significant for three items ("feeling low", "dizziness", and "afraid") and approaching significance for others. For the item "feeling low", boys showed an ICC which is lower than desired for psychometric purposes. For girls, however, this item yielded a coefficient in the upper range, suggesting gender differences in interpretation or experience of this symptom. The timing of puberty may be one explanation of the noted gender differences as puberty was found to influence the level of health complaints studied (Aro and Taipale, 1987). Prevalence of complaints may be another factor that contributes to gender differences. Otter *et al.* (1995) found that the interpretation needed to formulate an answer must be easily accessible in the respondent's memory to produce reliable test–retest data. Epidemiological studies showed that girls reported more health complaints than boys did (Belmaker, 1984; Aro and Taipale, 1987; King *et al.*, 1996; Klepp *et al.*, 1996). Thus, familiarity with the health complaints studied may have led to more reliable answers for girls. On the other hand, gender differences in reliability may produce differences in the level of reported complaints, and contribute to explaining the higher level of symptoms reported for girls.

Limitations of the study

In the present study, adolescents who were absent on the day of either test or retest did not take part. Previous studies found that nearly half of the children with widespread pain reported school absence because of pain (Mikkelsen *et al.*, 1997). This may have biased the results as adolescents who experience complaints often and of high intensity may have been excluded. A potential weakness of this study is the relatively short test–retest interval as memory may have influenced the results. However, long test–retest intervals may lead to

changes in the phenomenon under study and underestimate reliability. The substantial inter-item differences found in the present study do not suggest a general memory effect. Written consent from parents was a criterion of inclusion in the present study. Previous studies suggest that children of non-respondents may differ from the sample studied (Klepp, 1995). The latter consideration and a non-randomized sampling procedure may represent threats to validity and limit the possibility to generalize the results.

Conclusion

This study suggests that adolescents are able to understand, evaluate and report subjective health complaints. Most of the items in the studied checklist were found to have adequate validity. All items showed satisfactory test–retest reliability in the total sample, but with noticeable inter-item differences. Most changes were within one category. Girls generally showed higher levels of stability both for the total score and most single items. A few items showed ambiguity when described in interviews in spite of adequate test–retest stability. The findings confirmed the need to test validity in the target population and suggest that gender differences should be a focus in future studies of adolescent health measurement.

Acknowledgements

The Research Council of Norway provided funding for this paper. We thank Therese Andrews and Torbjørn Torsheim for commenting on an earlier version of this paper, and the latter for kindly preparing the data-files for the test–retest. We also thank the participating adolescents for sharing their information.

References

- Aaroe, L. E. and Wold, B. (1993). *Health behaviour in school aged children*. A WHO cross-national survey. Research protocol for the 1985–86 study. University of Bergen, Norway, 1986.
- Alfven, G. (1993). The covariation of common psychosomatic symptoms among children from socioeconomically different residential areas. An epidemiological study. *Acta Paediatrica*, **82**, 484–487.
- Aro, H. and Taipale, V. (1987). The impact of timing of puberty on psychosomatic symptoms among fourteen to sixteen-year-old Finnish girls. *Child Development*, **58**, 261–268.
- Belmaker, E. (1984). Nonspecific somatic symptoms in early adolescent girls. *Journal of Adolescent Health Care*, **5**, 30–33.
- Bendelow, G. A. and Williams, S. J. (1995). Transcending the dualism: towards a sociology of pain. *Sociology of health and illness*, **17**, 139–165.
- Burgher, M. S., Rasmussen, V. B. and Rivett, D. (1999). *The European Network of Health Promoting Schools. The alliance of education and health*. Copenhagen, WHO-Euro.
- Cioffi, D. (1991). Beyond attentional strategies: a cognitive-perceptual model of somatic interpretation. *Psychological Bulletin*, **109**, 25–41.
- Cioffi, D. (1996). Making public the private: Possible effects of expressing somatic experience. *Psychology and Health*, **11**, 203–222.
- Cronbach, L. (1971). Test validation. In Thorndike, R. (Ed.) *Educational Measurement*, 2nd ed, Washington, D.C.: American Council of Education.

- Eiser, C., Havermans, T. and Eiser, J. R. (1995). The emergence during adolescence of gender differences in symptom reporting. *Journal of Adolescence*, **18**, 307–316.
- Fallon, T Jr and Schwab-Stone, M. (1994). Determinants of reliability in psychiatric surveys of children aged 6–12. *Journal of Child Psychology and Psychiatry*, **35**, 1391–1408.
- Ferguson, G. A. and Takane, Y. (1989). *Statistical Analysis in Psychology and Education*. International editions. Psychology series. New York: Mc Graw-Hill Book company.
- Garralda, M. (1992). A selective review of child psychiatric syndromes with a somatic presentation. *British Journal of Psychiatry*, **61**, 759–773.
- Goodman, J. E. and McGrath, P. (1991). The epidemiology of pain in children and adolescents: a review. *Pain*, **46**, 247–264.
- Haugland, S., Wold, B., Stevenson, J., Aaroe, L. E. and Woynarowska, B. (2001). Subjective health complaints in adolescence—a cross-national comparison of prevalence and dimensionality. *European Journal of Public Health*, **11**, 4–10.
- King, A., Wold, B., Tudor-Smith, C. and Harel, Y. (1996). *The health of youth. A Cross-National Survey*. Canada: WHO Regional Publications. European Series No 69.
- Klepp, K. I. (1995). Nonresponse bias due to consent procedures in school-based, health-related research. *Scandinavian Journal of Social Medicine*, **23**, 53–59.
- Klepp, K. I., Aas, H., Maeland, J. G. and Alsaker, F. (1996). Self-reported health problems among school-students: a three-year follow-up study. *Journal of the Norwegian Medical Association*, **116**, 2032–2037 (in Norwegian).
- Kolip, P. and Hurrelmann, K. (1995). Coping strategies of Dutch and German adolescence. Psychosomatic complaints, aggressive behaviour, and the consumption of legal and illegal drugs. In *Childhood and youth in Germany and The Netherlands. Transitions and coping strategies of adolescents*, Du Bois-Reymond, M., Diekstra, R., Hurrelmann, K. and Peters, E. (Eds). Berlin: Walter de Gruyter, 183–197.
- Kujala, U. M., Taimela, S. and Viljanen, T. (1999). Leisure time physical activity and various pain symptoms among adolescents. *British Journal of Sports Medicine*, **33**, 325–328.
- Mechanic, D. (1983). Adolescent health and illness behaviour. *Journal of Human Stress* **9**, 4–13.
- Mikkelsen, M., Salminen, J. and Kautiainen, H. (1997). Non-specific musculoskeletal pain in preadolescents. Prevalence and 1-year persistence. *Pain*, **73**, 29–35.
- Mikkelsen, M., Salminen, J. and Kautiainen, H. (1996). Joint hypermobility is not a contributing factor to musculoskeletal pain in preadolescents. *Journal of Rheumatology*, **23**, 1963–1967.
- Millstein, S. G. (1993). A view of health from the adolescent's perspective. In Millstein S. G., Petersen, A. and Nightingale, E. O. (Eds). *Promoting the health of adolescents*, New York: Oxford University Press, 97–118.
- Otter, E. M., Mellenbergh, G. J. and deGlopper, K. (1995). The information between relation processing variables and test-retest stability for questionnaire items. *Journal of Educational Measurement*, **32**, 199–216.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Thousand Oaks, CA: Sage.
- Pennebaker, J. W., Burnam, M. A., Schaeffer, M. A. and Harper, D. (1977). Lack of control as a determinant of perceived physical symptoms. *Journal of Personality and Social Psychology*, **35**, 165–174.
- Pennebaker, J. W. (1982). *The psychology of physical symptoms*. New York, Heidelberg Berlin: Springer-Verlag.
- Pennebaker, J. W. and Roberts, T. A. (1992). Towards a his and hers theory of emotion: gender differences in visceral perception. *Journal of Social and Clinical Psychology*, **11**, 199–212.
- Shrout, P. E. and Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, **86**, 420–428.
- Starfield, B., Katz, H., Gabriel, A., Livingston, G., Benson, P., Hankin, J., Horn, S., Steinwachs, D. (1984). Morbidity in childhood—a longitudinal view. *New England Journal of Medicine*, **310**, 824–829.
- Starfield, B., Bergner, M. and Ensminger, M. *et al.* (1993). Adolescent health status measurement: development of the child health and illness profile. *Pediatrics*, **91**, 430–435.
- Streiner, D. L. and Norman, G. R. (1995). *Health Measurement Scales. A Practical Guide to their Development and Use*. Second edition. Oxford: Oxford University Press.

- Tellnes, G., Svendsen, K. O., Bruusgaard, D. and Bjerkedal, T. (1989). Incidence of sickness certification. *Scandinavian Journal of Primary Health Care*, **7**, 11–17.
- vanWijk, C. M. T. and Kolk, A. M. (1997). Sex differences in physical symptoms: the contribution of symptom perception theory. *Social Science and Medicine*, **45**, 231–246.
- vanWijk, C. M. T., Huisman, H. and Kolk, A. M. (1999). Gender differences in physical symptoms and illness behaviour—A health diary study. *Social Science and Medicine*, **49**, 1061–1074.
- Wisniewski, J., Naglieri, J. and Mulick, J. (1988). Psychometric properties of a children's psychosomatic checklist. *Journal of Behavioral Medicine*, **11**, 497–507.
- Wool, C. A. and Barsky, A. J. (1994). Do women somatize more than men? Gender differences in somatization. *Psychosomatics*, **35**, 445–452.