

# Paper pencil versus pc administered querying of a study on health behaviour in school-aged children

by

Vereecken C. <sup>1</sup>

---

## Abstract

**Purpose:** To investigate the reliability, the concordance, the feasibility and the appreciation of the computer format (pc format) of the Flemish version of the "Health Behaviour in School-aged Children-questionnaire" (HBSC-questionnaire) versus the original paper pencil format (pp format).

**Procedure:** Five hundred sixty first grade pupils of four secondary schools in Ghent completed the HBSC-questionnaire twice, with a test-retest interval of 6-10 days under one of the following conditions:

- twice the pc format (N = 177),
- twice the paper pencil format (N = 224),
- or both formats in counterbalanced order (N = 159).

**Results:** Test-retest reliabilities of both formats were at best moderate but comparable. For the majority of items mode of administration had no significant effect on adolescents' responses, but the pc format showed fewer item incompletions. Most of the pupils who completed both formats preferred the pc format.

---

Author responsible for correspondence: Carine Vereecken, Buffelstraat 17 – 9240 ZELE, Belgium – Fax: 09/240 49 94 – e-mail: Carine.Vereecken@rug.ac.be

<sup>1</sup> Ghent University, Department of Public Health, De Pintelaan 185 – 9000 Ghent, Belgium.

**Conclusion:** *The computer assisted version of the hbsc-questionnaire can be used as a valid alternative for the original paper and pencil version.*

## Keywords

Computer-applications, Health behaviour, Survey, School pupils.

## Introduction

The "Health Behaviour in School-aged Children" (HBSC) -study is a cross-national survey supported by the World Health Organization (Europe). The overall goal of the study is to gain new insights into – and to increase understanding of – health behaviour, lifestyles and their context in young people. Questionnaires have been chosen as the standard instrument for the collection of data. The international standard version of the questionnaire for each survey has been developed through cooperation between countries, with consensus reached through international planning meetings. A detailed description of the aims and theoretical framework of the study has been described elsewhere (1).

The international surveys are carried out every four years; in some countries national surveys are done every two years. Flanders has data for 1990, 1994, 1996 and 1998.

The populations selected for sampling are all school pupils aged 11, 13 and 15 years. The recommended minimum sample size for each of the three age groups is 1536 (1). "Cluster sampling" is used for sampling, where the cluster is the classroom (1). The questionnaires are completed in the school classroom and confidentiality is ensured.

Until now, school facilities did not allow systematic collection of data through PC administered questionnaires in the school.

However, in 1998 the Flemish government started an action program "PC/KD" to stimulate the use of computers and educational software at school. The goal is to have one PC per 10 pupils in every Flemish school by the end of 2002, starting from the 4th grade primary school education.

Because the cost of paper and pencil administration, the time required to produce a data set of 12000 questionnaires (the study sam-

ple in 1998), and the growing use of computers in school, a computer-administered format of the HBSC-questionnaire (pc format) might be an alternative form of the original paper pencil format (pp format).

The electronic survey eliminates the need for a person or technology to stand between respondents and the computer that stores and analyses their data (2). Some open-ended questions would still have to be coded, but sorting the data and making use of search engines can accelerate this process.

The instrument can be standardized easily (as is true for a pp format), while it still retains a lot of his flexibility (the questionnaire can be directed through a specific pathway, it can include error corrections and prompts, unnecessary questions can be skipped, it can handle open-ended questions requiring long answers, etc.) (2, 3).

However, recent studies suggest that while computer administered questioning has a place in the survey process, this technology may introduce response effects that would not be found in traditional data collection methods (4): it can encourage more guessing to multiple response questions (5), more positive agreement to scale attitude questions (6), more cooperation and willingness to participate (7, 8), fewer item incompletions and mistakes (2), increased reporting of some sensitive behaviours (9), or less social desirability in responses (2, 10) and greater likelihood to use the extreme response categories (4).

Thus, systematic comparisons between the pc format and the pp format of a questionnaire must be conducted, before the pc format can be considered as a valid alternative of the written questionnaire.

To date, only two studies have systematically compared computer-assisted versus written methods for anonymously collected data on a wide range of specific adolescent health behaviour (9, 11).

In the study of Turner et al. (9) an audio computer-assisted self interviewing (audio-CASI) technology for measuring sensitive behaviours (HIV-risk behaviours, drug use and interpersonal violence) was compared with a more traditional self-administered questionnaire. The results indicated that respondents were more likely to report sensitive behaviours when they were interviewed with the audio-CASI technology than when they were interviewed with the more traditional paper format.

In the study of Webb et al (11) health protective behaviours (seat belt use, physical activity and attention paid to eating habits), substance use

and sexual behaviours were measured by either pc format or pp format. Across the majority of health behaviours, mode of administration had no significant effect on adolescents' responses.

Both studies compared respondents completing either computer assisted questionnaires or written questionnaires. Neither of them compared the two methods with each subject acting as his or her own control.

## **Method**

### *Participants*

Participants were 657 male and female first grade pupils of four secondary schools in the neighbourhood of Ghent. In three schools at least 20 computers were available for querying the pupils, in the fourth school only 12 computers were available. All types of education were represented.

School classes were randomly assigned to one of three test scenarios.

- Group A was tested twice by computer.
- Group B served as a control group: participants were tested twice with the paper-pencil format.
- Group C: half of them were first tested with the paper-pencil format and retested by computer and half of them were first tested by computer and retested by the paper pencil format.

If for a certain class group not enough pc's were available, as many pupils as possible were given the pc format and the remainder of the pupils were given the pp format.

All pupils were told that the purpose of the study was to evaluate the test and not them as individuals.

## **Material**

A pilot version of the Flemish HBSC-questionnaire was administered. Survey items measured demographics and health behaviours on several specific domains: health protective behaviours, risk behaviours, food consumption, psychological and physical health, school, leisure activity and social relations.

The paper format was the usual HBSC-questionnaire.

The computer format was developed through the utilization of visual basic 4.0.

Test items were answered by responding with the mouse on a command button or choice field, or by typing the answer in a text field.

A skip button "I don't want to answer this question" was available for each question, except for the key variables age and sex. Answering with this button was, however, discouraged by a message box giving the standard instruction to minimize the use of this button, shown the 5<sup>th</sup>, 9<sup>th</sup>, 12<sup>th</sup>, 14<sup>th</sup> and further, each time this button was used.

For open ended questions an extra "next question" button was provided. If no answer was given, a message box appeared with a neutral statement indicating that there was still an answer expected and the program returned to the unanswered question.

An initial empty bar on the right side of the screen that fills in every time a question has been answered, showed how much of the questionnaire was done.

For a few questions a verification system was incorporated in the program <sup>1</sup>. If the respondent gave contradictive answers for two questions regarding the same topic, a confirmation regarding the last answer was asked and, if necessary, the previous question was displayed. If an incorrect answer was clicked on a second time, the incorrect response was stored.

Back paging to previous questions was possible by pushing a button labelled "previous question": at the same time the previous answer was removed.

As the last question was answered, the program was terminated and there was no possibility for the respondents to edit their answers.

---

<sup>1</sup> Items used for verification:

- sports-times per week vs sports-hours per week
- tried alcohol vs consumption frequency of beer, wine and spirits vs been drunk
- consumption frequency of the food items vs 24 hour recall of these items
- lives with grandparents/other people: if more than 4 grandparents or other people were living with them

## Procedure

The questionnaires were completed anonymously in a class setting and in the presence of one researcher and one teacher. The students were verbally instructed by the researcher how to complete the questionnaire. One period (50 min) was provided to complete the questionnaire.

Participants were not told that they would complete identical questionnaires twice. The time interval between testing and retesting ranged from 6 to 10 days.

After participants completed the questionnaire, their attitude toward the given format was assessed.

Participants who completed both formats were asked to indicate their preference.

To match test and retest, students were given a number that they had to adhere in their school diary.

## Statistics

Test-retest reliability as well as concordance between both formats were investigated. Reliability concerns the extent to which a measuring procedure yields the same results on independent repeated trials under the same conditions (12). Concordance between two variables shows the extent to which one of the variables can serve as a surrogate of the other (13).

Test-retest reliability, as well as concordance between two formats may be estimated by various statistical techniques. Kappa statistics were computed for variables at the categorical level, weighted kappa's for variables at the ordinal level and intraclass correlation coefficients for variables at the interval level.

Weighted Kappa values were computed and compared as described by Kramer and Feinstein (13).

The intraclass correlation coefficients (ICC), were computed as described by Deyo et al. (14).

The stability and the concordance were assessed for separate items, as well as for mean scores of items that were part of a scale.

The Wilcoxon signed rank test was used to measure systematic differences between both measurement methods for the ordinal variables. The McNemar test was used to measure systematic differences for dichotomous variables.

The mean number of end points selected over all questions that consisted of at least 5 answering categories (MEPS) was computed. A 2 x 2 x 2 repeated measures anova, with within subject factor "testing session" (first and second measurement) and between subject factors "method used during first session" (pc or pp), and "same or different method used during second session" (same or different) was used to investigate a possible time, method or order effect on the interval variable MEPS.

Since we were interested in patterns of findings across specific adolescent health domains, items were grouped according to these domains. Since a large number of coefficients were computed, attention was focused on the general pattern rather than on any specific coefficient.

## Results

### *Sample*

Five hundred sixty students had duplicate questionnaires (85.2%) (pc-pc: 177; pp-pp: 224; pc-pp: 73; pp-pc 86). Reasons for drop out were: absent matching code (one class dropped out completely due to practical problems (4%), illness on day of testing (4%), incorrect identification number (3.1%)), incomplete or not seriously completed questionnaires (pp format = 2.6%; pc format = 0.9%) and a program error (0.2%).

Due to practical problems (teachers forgetting the appointment, students forgetting their pencils...) some time was lost in several classes and not all students reached the end of the questionnaire in the remaining time interval. These variables were given a specific code so that it would be possible to distinguish them from real missings. In group A 126 pupils reached the last question at both occasions, in group B 181 pupils, and in group C 97.

Mean age of the participants was 13.6 years (st. dev. 0.79); 60% of the participants were boys. In group C 49% of the participants were boys.

In the other two groups subject gender was balanced as much as possible (67% were boys in group A, 63% in group B). In group A 31% of the respondents followed vocational oriented education, in group B 13% and in group C 8%.

### *Test retest stability*

As shown in table 1 to 7, test-retest kappa statistics ranged from 0.31 to 0.99 for the pc format (group A) and from 0.11 to 1 for the pp format (group B). Mean kappa of the pc format was 0.59 (st. dev. 0.16) and of the pp format 0.62 (st. dev. 0.15).

Comparing kappas of the pp format with those of the pc format, showed that differences were small but in favour of the pp format: kappas of the pp format were 0.10 or more points higher than those of the pc format in 17 comparisons while kappas for the pc format were 0.10 or more points higher than those of the pp format in only 5 comparisons. Statistical analyses revealed, however, only four significant differences at the 0.05 level [helpless ( $k_{pc-pc}$  0.31;  $k_{pp-pp}$  0.59;  $z = -2.92$ ), feel lonely ( $k_{pc-pc}$  0.32;  $k_{pp-pp}$  0.61;  $z = -2.85$ ), parents expect too much ( $k_{pc-pc}$  0.39;  $k_{pp-pp}$  0.60;  $z = -2.18$ ) and times with friends after school ( $k_{pc-pc}$  0.44;  $k_{pp-pp}$  0.63;  $z = -2.11$ )] all four in favour of the pp format.

For the quantitative data (age, amount of pocket money and number of cigarettes/week) and for the scale scores (school atmosphere, teacher, student and parent relations), ICC ranged between 0.42 and 0.91 for the pc format and between 0.27 and 0.81 for the pp format.

Test retest stability did not differ for those items that incorporated the verification system in the pc format

### *Concordance*

For calculating the agreement between both formats, data of pc and pp format of group C were used. Kappa statistics ranged from 0.08 to 1 (table 1 to 7: last column), with a mean kappa statistic of 0.60 (st. dev. 0.15). The ICC ranged from 0.51 to 0.95. Kappa's and ICC's were comparable with those of the test retest formats.

### *Systematic differences involving mode of administration*

Looking at systematic differences between test retest of the same format, 17 significant differences were found for the pc format and 19 sig-



TABLE 1  
Kappa statistics and ICC of demographics

			PC-PC	PP-PP	PP-PC
sex	(1 = boy; 2 = girl)	κ	0,99	1,00	1,00
age		ICC	0,84	0,81	0,94
lives with (steph) mother	(1 = mother; 2 = steph; 3 = both; 4 = none)	κ	0,61	0,76	0,59
lives with (steph) father	(1 = father; 2 = steph; 3 = both; 4 = none)	κ	0,83	0,92	0,69
lives with sisters	(0 = none; 8 = more than 7)	κ	0,91	0,91	0,91
lives with brothers	(0 = none; 8 = more than 7)	κ	0,94	0,93	0,86
lives with grandparents	(0 = none; 8 = more than 7)	κ	0,47	0,49	0,28
lives with other people	(0 = none; 8 = more than 7)	κ	0,38	0,45	0,08
SES father	(1 = high; 5 = low)	κ	0,92	0,88	0,84
SES mother	(1 = high; 5 = low)	κ	0,95	0,86	0,87
pocket money	(1 = yes; 2 = no)	κ	0,76	0,85	0,75
amount of pocket money		ICC	0,91	0,63	0,95
family well off	(1 = very well off; 5 = not at all well of)	κ	0,67	0,72	0,60
family car	(1 = no; 3 = two or more)	κ	0,79	0,89	0,81
own bedroom	(1 = yes; 2 = no)	κ	0,88	0,92	0,81
holidays	(1 = not at all; 4 = more than twice)	κ	0,59	0,70	0,74

TABLE 2  
Kappa statistics and ICC's of school variables

			PC-PC	PP-PP	PP-PC
academic achievement	(1 = very good; 4 = below average)	κ	0,70	0,74	0,81
liking school	(1 = like a lot; 4 = not at all)	κ	0,60	0,75	0,73
parent expect too much	(1 = strongly agree; 5 = strongly disagree)	κ	0,39	0,60	0,62
teacher expect too much	(1 = strongly agree; 5 = strongly disagree)	κ	0,38	0,47	0,49
school is boring	(1 = very often; 5 = never)	κ	0,52	0,55	0,52
alone at school	(1 = hasn't happened; 5 = several times a week)	κ	0,38	0,43	0,40
feel safe at school	(1 = always; 5 = never)	κ	0,42	0,49	0,41
pressured by school work	(1 = not at all; 4 = a lot)	κ	0,42	0,44	0,40
school atmosphere	(5 items: 1 = strongly agree; 5 = strongly disagree)	ICC	0,58	0,61	0,78
teacher relations	(4 items: 1 = strongly agree; 5 = strongly disagree)	ICC	0,48	0,63	0,61
student relations	(3 items: 1 = always; 5 = never)	ICC	0,42	0,75	0,69
parent relations	(3 items: 1 = always; 5 = never)	ICC	0,60	0,66	0,51
been bullied	(1 = haven't; 5 = several times a week)	κ	0,42	0,52	0,55
have bullied	(1 = haven't; 5 = several times a week)	κ	0,48	0,65	0,60
truancy	(1 = 0 days; 5 = four days or more)	κ	0,42	0,60	0,46

TABLE 3  
Kappa statistics of variables relating to psychological and physical health

		PC-PC	PP-PP	PP-PC
reported health	(1 = very healthy; 3 = not very healthy)	0,53	0,59	0,60
quality of life	(1 = very happy; 4 = not happy at all)	0,53	0,60	0,52
feel lonely	(1 = very often; 4 = no)	0,32	0,61	0,56
left out of things	(1 = always; 5 = never)	0,52	0,56	0,50
helpless	(1 = always; 5 = never)	0,31	0,59	0,54
feel confident	(1 = always; 5 = never)	0,40	0,35	0,44
headache	(1 = about every day; 5 = rarely or never)	0,51	0,60	0,52
stomachache	(1 = about every day; 5 = rarely or never)	0,45	0,52	0,48
backache	(1 = about every day; 5 = rarely or never)	0,56	0,45	0,55
feeling low	(1 = about every day; 5 = rarely or never)	0,44	0,57	0,51
irritable-bad temper	(1 = about every day; 5 = rarely or never)	0,35	0,41	0,43
nervous	(1 = about every day; 5 = rarely or never)	0,41	0,49	0,45
sleeping difficulties	(1 = about every day; 5 = rarely or never)	0,53	0,45	0,48
dizziness	(1 = about every day; 5 = rarely or never)	0,47	0,50	0,48
tired in the morning	(1 = rarely or never; 4 = 4 or more times a week)	0,48	0,59	0,58
use of medicine-headache	(1 = no; 3 = more than once)	0,53	0,57	0,52
use of med.-stomachache	(1 = no; 3 = more than once)	0,46	0,64	0,47
use of med.-sleep. Difficulties	(1 = no; 3 = more than once)	0,57	0,11	0,41
use of med.-nervousness	(1 = no; 3 = more than once)	0,65	0,41	0,48
change body	(1 = yes; 2 = no)	0,59	0,68	0,69
think about body	(1 = much too thin; 5 = much too fat)	0,63	0,67	0,71
think about looks	(1 = very good looking; 5 = not at all)	0,53	0,57	0,66

nificant differences for the pp format (data not shown). Of these differences, seven were in common for pc and pp format.

No systematic pattern could be revealed in these differences: some items were answered more positive during the retest while others were answered more negative.

Comparing systematic differences between pc and pp revealed 10 significant differences (table 8). The sign of the difference is indicated by

TABLE 4  
Kappa statistics and ICC of risk behaviours

			PC-PC	PP-PP	PP-PC
tried smoking	(1 = yes; 2 = no)	κ	0,85	0,87	0,75
smoking frequency	(1 = every day; 4 = do not smoke)	κ	0,89	0,83	0,85
no cigarettes		ICC	0,86	0,27	0,94
use of hash	(1 = never; 3 = 3 times or more)	κ	0,75	0,39	0,55
tried alcohol	(1 = yes; 3 = no)	κ	0,75	0,73	0,71
Beer	(1 = every day; 5 = never)	κ	0,70	0,65	0,64
Wine	(1 = every day; 5 = never)	κ	0,64	0,59	0,63
Spirits	(1 = every day; 5 = never)	κ	0,53	0,57	0,55
been drunk	(1 = never; 5 = more than 10 times)	κ	0,80	0,81	0,75

TABLE 5  
Kappa statistics of health protective behaviours

		PC-PC	PP-PP	PP-PC
tooth brushing	(1 = never; 7 = more than once every day)	0,70	0,80	0,75
use of seatbelt	(1 = always; 4 = rarely or never)	0,82	0,81	0,73
use of bike helmet	(1 = always; 4 = rarely or never)	0,59	0,55	0,69

TABLE 6  
Kappa statistics of nutrition related variables

		PC-PC	PP-PP	PP-PC
Consumption frequency				
Fruit	(1 = never; 7 = more than once every day)	0,54	0,58	0,54
Vegetables	(1 = never; 7 = more than once every day)	0,50	0,45	0,53
Sweets	(1 = never; 7 = more than once every day)	0,44	0,50	0,43
regular soft drinks	(1 = never; 7 = more than once every day)	0,59	0,60	0,57
(semi-)skimmed milk	(1 = never; 7 = more than once every day)	0,60	0,60	0,62
whole fat milk	(1 = never; 7 = more than once every day)	0,61	0,55	0,67
brown bread	(1 = never; 7 = more than once every day)	0,72	0,58	0,69
Crisps	(1 = never; 7 = more than once every day)	0,49	0,52	0,49
Chips	(1 = never; 7 = more than once every day)	0,49	0,40	0,50
Coffee	(1 = never; 7 = more than once every day)	0,83	0,74	0,75
on a diet	(1 = no weight = fine; 2 = no, need to; 3 = no to skinny; 4 = yes)	0,81	0,81	0,79

TABLE 7  
*Kappa statistics of variables relating to social relations and leisure activity*

		PC-PC	PP-PP	PP-PC
easy to talk to father	(1 = very easy; 4 = very difficult)	0,59	0,60	0,59
easy to talk to mother	(1 = very easy; 4 = very difficult)	0,69	0,62	0,60
easy to talk to elder sister	(1 = very easy; 4 = very difficult)	0,51	0,54	0,59
easy to talk to elder brother	(1 = very easy; 4 = very difficult)	0,54	0,60	0,48
easy to talk to friends same sex	(1 = very easy; 4 = very difficult)	0,57	0,53	0,53
easy to talk to friends oppos sex	(1 = very easy; 4 = very difficult)	0,56	0,53	0,38
close friends	(1 = none; 4 = three or more)	0,55	0,66	0,64
easy to make friends	(1 = very easy; 4 = very difficult)	0,53	0,62	0,64
times with friends after school	(1 = 4-5 days a week; 4 = no friends)	0,44	0,63	0,62
evenings with friends	(0,,7)	0,68	0,61	0,67
hours a day watch tv	(1 = not at all; 6 = more than 4)	0,57	0,69	0,70
hours a week play PC games	(1 = not at all; 6 = 10 or more)	0,61	0,69	0,65
sports-times per week	(1 = every day; 7 = never)	0,61	0,59	0,62
sports-hours per week	(1 = none; 6 = 7 or more)	0,53	0,61	0,52

the mean of the difference between the pp format and the pc format ( $\Delta(\text{pp-pc})$ ).

More pupils indicated to have tried smoking, to drink spirits and soft drinks frequently, to be tired in the morning, to use headache-and stomach-ache-medication, to skip school and to feel left alone at school in the pc format.

They indicated to find it easier to talk to friends of opposite sex and reported less grandparents living with them.

No differences were found in comparing the scores of the school attitude scales.

Of the 14 items incorporating the verification system, only two items showed a significant swift, namely the amount of grandparents living with them and the frequency of soft drink consumption.

#### *Amount of missings*

Comparing the mean amount of missings of the first inquiry, between both formats, showed that respondents of the pp format left 5.38% of the queried items blank, whereas this was only 1.34% of the pc format.

TABLE 8  
Mean (and standard error) of the difference between the pp format and the pc format ( $\Delta(pp-pc)$ ) and significance of the test for symmetry

		mean(se)	p
lives with grandparents alone at school	(0 = none; 8 = more than 7)	0,19(0,06)	0,00
Truancy	(1 = hasn't happened; 5 = several times a week)	-0,21(0,10)	0,04
tired in the morning	(1 = 0 days; 5 = four days or more)	-0,16(0,06)	0,01
use of medicine-headache	(1 = rarely or never; 4 = 4 or more times a week)	-0,16(0,07)	0,02
use of med.-stomachache	(1 = no; 3 = more than once)	-0,14(0,05)	0,01
tried smoking	(1 = no; 3 = more than once)	-0,09(0,04)	0,03
Spirits	(1 = yes; 2 = no)	0,06(0,03)	0,05
regular soft drinks	(1 = every day; 5 = never)	0,11(0,05)	0,03
easy to talk to friends oppos sex	(1 = never; 7 = more than once every day)	-0,31(0,12)	0,01
	(1 = very easy; 4 = very difficult)	0,22(0,09)	0,02

TABLE 9  
Average number of end points selected

	first session		second session	
	mean	st dev	mean	st dev
pp pp	0.389	0.125	0.397	0.135
pc pc	0.403	0.132	0.419	0.139
pp pc	0.397	0.128	0.402	0.136
pc pp	0.405	0.127	0.403	0.116

### *Tendency to use extreme scale points*

The tendency to use extreme scale points was calculated by computing the MEPS. Means and standard deviations of the MEPS are presented in table 9.

A 2 x 2 x 2 repeated measures anova, with within subject factor "testing session" (first and second measurement) and between subject factors "method used during first session" (pc or pp), and "same or different method used during second session" (same or different) showed no main effects nor interaction effects (table 10).

### *Feasibility and attitude toward both formats*

Forcing respondents who administered both formats (group C) to choose between pp and pc, 94% selected the pc format as their preferred administration mode.

TABLE 10  
*Repeated measures anova on the mean number of end points selected*

Source of Variation	SS	DF	MS	F	P
Between subjects effect					
within cells	15.97	555	0.03		
same-different	0	1	0	0	0.979
pcpp	0.03	1	0.03	1.03	0.310
same-different by pcpp	0.01	1	0.01	0.39	0.530
Within subject effect					
within cells	3.04	555	0.01		
session	0.01	1	0.01	1.6	0.207
same-different by session	0.01	1	0.01	0.95	0.331
pcpp by session	0	1	0	0.01	0.923
same-different by pcpp by session	0	1	0	0.64	0.423

Reasons for this general preference were: like it more (35%), easier (32%), quicker (27%), less writing (13%), like to work on a pc (7%), you don't have to fill in bubbles (2.2%), don't know (1.5%), clearer (1.5%), an occasion to learn to work on a pc (1.5%), an occasion to work on a pc (1.5%), easier to correct mistakes (1.5%), impossible to forget to fill in questions (0.7%), less boring (0.7%) and more modern (0.7%).

Of the eight respondents who preferred the paper format, three viewed it as quicker, two didn't know why, one didn't like to work on a pc, one found it less tiring for his/her eyes and one missed the drawings of the paper pencil format.

A Mann-Whitney U test on the data of the first inquiry, revealed statistically significant differences ( $p < 0.001$ ) between pc and pp format on both 5 point semantic-differential items "difficult-easy" and "annoying-nice", in favour of the computer format (mean: 3.9 versus 3.6 and 3.6 versus 3.1). As well pupils of the more general oriented education levels as pupils of vocational level were able to complete the pc format of the questionnaire. Comparing respondents with limited pc experience (less than 10 times) with those who had more pc experience gave no significant differences.

## Discussion

As the HBSC-study is an international study, data can be compared between different countries and as the HBSC-study uses the same ques-

tions and answering categories every four years (in some countries every two years), trends can be computed. However, once the wording or mode of administration is changed, the data can change and trend analyses or comparability across administrations can be affected. Consequently, before introducing the pc format of the HBSC-questionnaire, empirical verification of the agreement between both formats was necessary.

### *Differences and agreements between both formats*

Because of mechanical differences between the formats, the instructions for the pc format were modified slightly (e.g. filling the appropriate bubble on the paper survey versus responding with the mouse on the appropriate button) and some questions showed minimal different wordings.

Additional, a verification system was incorporated for a few questions of the pc format. This system made it possible to point out to the respondents that they had misinterpreted a question and gave them the possibility to correct for it.

The progression bar in the pc version had to imitate the possible effect of seeing how much of the written questionnaire was completed.

The button "I don't want to answer this question", gave the respondents of the pc format the possibility to skip questions; respondents of the paper format who did not want to answer any specific questions could leave these blank. We believe that forcing respondents to answer questions, would heighten faulty – and therefore invalid – answers.

### *Test-retest*

Our results indicate that the test-retest reliabilities of the pc format and the pp format are in general at best moderate but comparable. Four significant different kappa's between pc and pp format were found, but comparing 83 times 2 kappa's, we could have expected that even by chance alone four kappa's differed significantly.

Looking at the ICC's, it strikes that open ended questions (amount of pocket money and cigarettes) are much less reliable for the pp format. We suspect that the unclearness of the pupils hand-writing plays an important part in this.

The tests for symmetry showed that, even without an intervention between test and retest, some shifts were found for the pc format as for the pp format. However, no general rule could be discovered in these shifts.

### *Response effects*

Cross-mode kappa statistics and ICC's were in line with the same-mode kappa statistics and ICC's.

Unlike the pp format, answers in the pc format immediately disappear when the next question is shown on the screen. This could mean that pupils are less embarrassed to report accurately in behaviours that are less accepted, when there is no possibility that a teacher or supervisor can look into their answers. Looking at systematic differences between both formats, the pc format resulted in higher frequencies of less socially desirable answers for eight items (more pupils indicated to have tried smoking, to drink spirits and soft drinks frequently, to be tired in the morning, to use headache- and stomach-ache-medication, to skip school and to feel left alone at school), suggesting minimization of social desirability bias (11). This conclusion assumes of course that the observed increase in reporting of these behaviours is more accurate, an assumption that follows standard practice in much social measurement research (9).

For the majority of items, we found however, that mode of administration had no significant effect on adolescents' responses, and this, as well, in the more sensitive as the more neutral domains.

We did not find more extreme answers for pupils using the pc format than for pupils using the pp format.

Like other studies, respondents preferred the computer format (10, 15, 16). Computer survey respondents were more enthusiastic (17, 18) and fewer items were left blank.

More pp questionnaires ( $n = 17$ ) than pc questionnaires ( $n = 6$ ) were invalid and had to be excluded from analyses. This difference could be caused by a more favourable attitude toward the pc format, but also because it is easier to discover frivolously completed written questionnaires than frivolously completed computer administered questionnaires: deviant answers were not feasible in the pc format and systematic answering patterns are easier recognized on a paper.



To compensate this shortcoming of the pc format, we recommend to include parallel questions (nearly impossible in a pp format because back paging is much more likely).

## **Conclusion**

The principal finding of this study is that the stability as well as the general accuracy of the pc format of the HBSC-questionnaire is comparable with the pp. So, we conclude that the pc format of the hbsc-questionnaire can be used as an alternative for the pp format, if extra attention is paid in analysis of those items showing a systematic difference.

However, as the present study was conducted in pupils of the first grade secondary school (13-14 year of age), and as there is little known about computer administered querying in younger children, further research is needed to establish whether younger pupils (11-12 year of age) have enough skills to complete the pc format of the HBSC-questionnaire.

## **Acknowledgment**

The author thanks Annelies Van Hove for her assistance with data collection.

## **Samenvatting**

**Doel:** De bruikbaarheid, de betrouwbaarheid, de overeenstemming en de waardering van een computergestuurde versie van de vragenlijst "Health Behaviour in School-aged Children" vergelijken met de originele gedrukte versie.

**Procedure:** Vijf honderd zestig eerste graad studenten van 4 secundaire scholen in de omgeving van Gent vulden de vragenlijst tweemaal in, met een test-hertest interval van 6-10 dagen, onder een van volgende condities:

- tweemaal de pc versie (N = 177),
- tweemaal de gedrukte versie (N = 224),
- of beide versies (N = 159) waarvan 73 eerst de computergestuurde versie en 86 eerst de gedrukte versie.

**Resultaten:** Test-hertest betrouwbaarheidscoëfficiënten van beide versies waren gematigd maar vergelijkbaar.

Voor de meerderheid van de items had de manier van testafname geen significant effect op de antwoorden van de respondenten, doch de computergestuurde versie vertoonde minder ontbrekende waarden.

De meeste leerlingen die beide versies invulden, verkozen de computergestuurde versie.

**Besluit:** De computergestuurde versie van deze vragenlijst kan als alternatief voor de gedrukte versie gebruikt worden.

## References

1. CURRIE C. Health Behaviour in School-Aged Children: Research Protocol for the 1997-1998 Survey. 1998.
2. KIESLER S, SPROULL L S. Response Effects in the Electronic Survey. *Public Opinion Quarterly* 1986; 50(3): 402-413.
3. DILLMAN D A. Mail and other self-administered questionnaires. In: Rossi P H, Wright J D, Anderson A B. *Handbook of Survey Research*. New York: Academic Press; 1983: 359-377.
4. BRATTON G R, NEWSTED P R. Response Effects and Computer-Administered Questionnaires: The Role of the Entry Task and Previous Computer Experience. *Behaviour and Information Technology* 1995; 14(5): 300-312.
5. LIEFELD J P. Field experience: computer interviewing with consumer samples. Synodinos N. *Proceedings of the American Marketing Association Microcomputer Workshop*. University of Hawaii: 1987: 129-140.
6. LIEFELD J P. Response Effects in Computer-Administered Questioning. *Journal of Marketing Research* 1988; 25(4): 405-409.
7. SPROULL L S. Using Electronic Mail for Data Collection in Organizational Research. *Academy of Management Journal* 1986; 29(1): 159-169.
8. LIEFELD J P. Computer interviewing: cursor and response task effects. *Canadian Journal of Marketing Research*. 1987; 6: 20-25.
9. TURNER C F, KU L, ROGERS S M, LINDBERG L D, PLECK J H, SONENSTEIN F L. Adolescent Sexual Behavior, Drug Use, and Violence: Increased Reporting With Computer Survey Technology [See Comments]. *Science* 8-5-1998; 280(5365): 867-873.
10. LOCKE S D, GILBERT B O. Method of Psychological Assessment, Self-Disclosure, and Experiential Differences: A Study of Computer, Questionnaire, and Interview. *Assessment Formats*. *Journal of Social Behavior and Personality* 1995; 10(1): 255-263.
11. WEBB P M, ZIMET G D, FORTENBERRY J D, AND BLYTHE M J. Comparability of a Computer-Assisted Versus Written Method for Collecting Health Behavior Information From Adolescent Patients. *J. Adolesc. Health* 1999; 24(6): 383-388.
12. CARMINES E G, ZELLER R A. *Reliability and Validity Assessment*. 1979. Beverly Hills, California. Sage Publications.
13. KRAMER M S, FEINSTEIN A R. *Clinical Biostatistics*. LIV. The Biostatistics of Concordance [Published Erratum Appears in *Clin Pharmacol Ther* 1989 Sep; 46(3): 309]. *Clin. Pharmacol. Ther.* 1981; 29(1): 111-123.
14. DEYO R A, DIEHR P, PATRICK D L. Reproducibility and Responsiveness of Health Status Measures. *Statistics and Strategies for Evaluation*. *Control Clin. Trials* 1991; 12(4 Suppl): 142S-158S.
15. ERDMAN H, KLEIN M H, GREIST J H. The Reliability of a Computer Interview for Drug Use/Abuse Information. *Behavior Research Methods, Instruments and Computers* 1983; 15(1): 66-68.

- 
16. WHITE D M, CLEMENTS C B, FOWLER R D. A Comparison of Computer Administration With Standard Administration of the MMPI. *Computers in Human Behavior* 1985; 1(2): 153-162.
  17. MILLSTEIN S G. Acceptability and Reliability of Sensitive Information Collected Via Computer Interview. *Educational and Psychological Measurement* 1987; 47(2): 523-533.
  18. SKINNER H A, ALLEN B A. Does the Computer Make a Difference? Computerized Versus Face-to-Face Versus Self-Report Assessment of Alcohol, Drug, and Tobacco Use. *Journal of Consulting and Clinical Psychology* 1983; 51(2): 267-275.