Journal of Official Statistics, Vol. 15, No. 3, 1999, pp. 413–429

A Disk by Mail Survey of Pupils in Primary Schools: Data Quality and Logistics

Marion J.C. van Hattum[1] *and Edith D. de Leeuw*[2]

Computer-assisted self-reporting and disk by mail (DBM) surveys are less widespread than other forms of computer-assisted data collection, and they are generally restricted to special populations. The Netherlands has the unique situation that each primary school has at least one PC, thanks to a government-sponsored project to enhance computer literacy. Thus, when we were faced with the challenge of a survey among young children on bullying in schools, we decided to implement a disk by mail survey.

Our main reason for using computer-assisted self-administered questionnaires (CSAQ) is the extremely sensitive nature of questions about bullying. Pupils are very reluctant to talk about bullying, even to their parents and teachers. CSAQ enhances the feeling of privacy, and in general does well with questions of a sensitive nature. In designing the questionnaires, we tried to make them as simple and attractive as possible.

Diskettes were sent to 106 Dutch primary schools for use on levels five, six, seven, and eight (pupils aged 8–12). The pupils were surveyed individually, using self-administered computer-assisted questionnaires. The questionnaire focussed on attitudes to bullying and actual bullying (either as victim or as culprit).

We were able to show that: (1) a DBM survey can be successfully implemented in Dutch primary schools; (2) children from the age of eight years can successfully complete a computer-assisted self-interview, and enjoy it; (3) the cited advantages of computer-assisted data collection (less item nonresponse, less social desirability) also hold when the subjects are children; (4) DBM results in less costs for each completed and returned questionnaire.

Key words: Bullying; young children; CSAQ; CASI; DBM; acceptance; item nonresponse; social desirability; costs.

1. Introduction

Traditionally, when surveying sensitive topics, researchers use self-administered questionnaires, either as a mail survey or as a paper questionnaire that is handed over by an interviewer. Self-administered questionnaires have the advantage that they evoke a greater sense of privacy and lead to more self-disclosure (Sudman and Bradburn 1974; Tourangeau and Smith 1996). Empirical research has shown that self-administered questionnaires when compared to interviews produce more valid reports of sensitive behavior and less socially desirable answers in general (e.g., Aquilino 1994; Hochstim

[1] Utrecht University, Trompstraat 96, NL-3814 St Amersfoort, The Netherlands. E-mail: mhattum@ trimbos.nL
[2] Plantage Doklaan 40, NL-1018 CN Amsterdam, The Netherlands. E-mail: edithL@educ.uva.nL

Acknowledgments: We especially thank all teachers and pupils for their enthusiasm and their cooperation during the data collection. We also thank Joop Hox and Sabina Kef for their support and the reviewers for their helpful comments. Special credit is due to Fons van Houte whose invaluable assistance made this project a success.

1967; Siemiatycki 1979; Turner, Lessler, and Devore 1992; for a comprehensive review see De Leeuw 1992). However, paper self-administered questionnaires have one serious drawback: only relatively simple questionnaires can be used (cf. Dillman 1978).

Complex paper self-administered questionnaires, with many skippings and branchings, negatively influence the survey quality in two ways. The length and complex structure enhance the perceived respondent burden and will lead to more explicit refusals to cooperate. When a respondent is willing to answer the questionnaire, the complexity increases the cognitive burden of the respondent, putting more strain on the question-answer process, which will negatively influence data quality (cf. Schwarz, Strack, Hippler, and Bishop 1991; Tourangeau and Smith 1996).

Computer Assisted Self-Administered Questionnaires (CSAQ) and Computer-Assisted Self-Interviewing (CASI) can overcome these problems, and make it possible to use very complex questionnaires without the aid of an interviewer. In these applications the interview program takes over and handles the questionnaire logic and question flow. Respondents simply read each question from the screen, type in an answer, and are no longer burdened with complex routing instructions[3]. Furthermore, the use of a computer may enhance the feeling of privacy. After an answer is given, it disappears from the screen, while an answer that is written down remains on the paper for everyone to see. Beckenbach (1995) showed in an experimental study that respondents evaluated sensitive questions as somewhat less threatening when CAPI and CSAQ were used than with a paper and pen interview. And in an extensive meta-analysis of 39 studies Weisband and Kiesler (1996) found that computer administration increases self-disclosure even when compared with paper-and-pencil (self-administered) forms.

Therefore, when designing a study on ''bullying in schools,'' we favored computer-assisted self-administered questionnaires. Admitting that one is being bullied is highly threatening a child's self-esteem, while admitting that one bullies will result in social disapproval or even punishment. Pupils are in general very reluctant to talk about bullying, even to their parents or teachers (Besag 1989; Heenan 1992). As the respondents are children, it is also important that the questionnaire *appears* simple and attractive. CSAQ can do this far better than a paper form, while simultaneously enhancing the feeling of privacy.

There are various forms of CSAQ or CASI (De Leeuw, Hox, and Snijkers 1995; Nicholls, Baker, and Martin 1997). We chose Disk by Mail for two reasons: first because this made it possible that pupils could use the trusted school computer they are acquainted with, and second because the costs of this type of survey are comparatively low.

For the successful implementation of Disk by Mail surveys it is necessary that the potential respondents have access to a personal computer, and that the researcher knows which types of computer-systems, monitors, and disks are used by the respondent (Saltzman 1993; Witt and Bernstein 1992; Zandan and Frost 1989). Thanks to a large government-sponsored project to improve computer literacy among the young, all primary

[3] Both the abbreviations CSAQ (Computer-Assisted Self-Administered Questionnaire) and CASI (Computer-Assisted Self-Interviewing) are used for those cases when the respondents *themselves* read the question from the screen and type in their answers. In general, the term CASI is used when an interviewer is present to introduce the survey and offer help when necessary, while the term CSAQ is used for those forms when no interviewer is involved, as is the case in Disk by Mail. We therefore will use the abbreviation CSAQ throughout this article.

schools in the Netherlands are equipped with personal computers[4]. Disks with two different questionnaires were sent to the schools: a questionnaire for pupils and a questionnaire for teachers. In this article we focus on the children's questionnaire. The majority of the schools used the computer-assisted questionnaires. A small subsample of schools used paper-and-pencil questionnaires.

The methodological aim of our study was threefold. First, we wanted to test the feasibility of computer-assisted data collection for pupils in primary schools. Although all schools used computers to some extent, schools and teachers did differ in computer experience. We are convinced that when implemented with care, DBM can be a success in primary schools. However, teachers who are older and less experienced with computers could have some problems. Also, the youngest pupils could pose problems. To test this and to evaluate the general acceptance of our method, we approached the teachers with a special paper questionnaire. This evaluation study was done after the data collection was completed at the schools.

The second aim was to determine if the use of CSAQ improves the quality of the resulting data, when *children* are surveyed. In the literature on computer-assisted data collection both the reduction of missing data and a larger extent of self-disclosure are well documented for adult populations (cf. Nicholls, Baker, and Martin 1997; Weisband and Kiesler 1996). An often cited disadvantage of computer-assisted data collection is the ''itemization'' of the questionnaire: questions appear one by one on the screen instead of grouped in meaningful clusters on pages (cf. De Leeuw, Hox, and Snijkers 1995; Nicholls et al. 1997). This itemization could disturb the concentration of the children, resulting in less consistency in the computer-assisted version. Using a quasi-experimental design, we evaluated the data of primary school children with regard to these three aspects.

Our third aim was to assess all costs, and compare the costs for a Disk by Mail survey of schools with the costs of a similar paper-and-pencil survey.

In the next sections, we first discuss the design of the study and the logistics of the data collection procedures. This is followed by the results of an evaluation survey of the procedures used. We then focus on a mode comparison, comparing the results of the CSAQ children's questionnaires with a smaller number of paper-and-pencil self-administered questionnaires. We also include a cost comparison for a Disk by Mail and paper-and-pencil mail survey.

2. Design

2.1. Population

The study took place in the groups five, six, seven and eight of Dutch primary schools. In the Netherlands, children start school in group one at the age of four, generally reach

[4] This project, called the Comenius project after the Czech educator Jan Amos Comenius (1592–1670), started in 1989. Each primary school in the Netherlands received several high-grade PCs with a hard disk, mouse, color monitors, Ms-Dos/Windows environment; every school also received a printer, basic software (text processing, drawing and educational software), and a small grant ($300) to buy additional software. The number of computers per school depended on the number of pupils: on average one computer was provided for every 60 pupils. Also, from each school at least one staff member participated in special courses, which meant that each school had certified ''computer activity coordinators'' among their teaching staff (cf. Van der Ree 1994). At present, these school computers are being upgraded and their number will be extended.

group five around the age of eight or nine, and generally leave primary school after group eight at the age of 12. Dutch school researchers start using paper questionnaires in group five, since by that time children have sufficient reading and writing skills to complete a questionnaire.

Standard procedure is that no subsampling takes place at school-level. Therefore, all children and their teachers were asked to complete a questionnaire.

2.2. *Sample and response*

In November 1994, 217 Dutch primary schools, which formed a random sample of Dutch primary schools, were contacted with a request to participate in a study on bullying. Initially 119 schools agreed to cooperate. One of the main reasons for refusal was that the teachers and the school administration were far too busy to spare the time. This is a general problem in school research and not related to the use of computers during the data collection. Other reasons mentioned were related to the content of the study. Some schools mentioned that they were not interested in a study about ''bullying'' because bullying was no problem at the school, or that they already had a special program to coach ''problem'' children. It should be emphasized that no specific mention of computers was made by the refusers. Of the 119 willing schools a total of 106 did complete the questionnaires in the spring of 1995. This brings the total response rate to 49%; for Dutch school research this is a very high response rate (average is 30–35%). The responding schools did not significantly differ from the nonresponding schools, nor from the total population of schools, on general school background characteristics such as denomination and pedagogical philosophy (cf. Van Hattum 1997).

2.3. *Quasi-experimental design*

In total 106 schools completed the survey. In 98 schools the data were collected by means of computer-assisted self-administered questionnaires. Eight schools used a paper-and-pencil version of these questionnaires. In total, 556 children completed a paper-and-pencil (PAPI) questionnaire about bullying and victimization, and 5,872 children a computer version. This enables us to compare results on both versions of the questionnaire and determine if the use of CSAQ improves the quality of the resulting data, when *children* are surveyed. Of course, schools were not randomly assigned to experimental conditions as in a full experimental (split sample) design. Instead a quasi-experimental design was used with existing groups (Cook and Campbell 1979). This quasi-experimental design has the major weakness that the groups may not be equivalent, for instance due to self-selection. We checked for equivalence in the following two ways.

First, we investigated the reasons for using PAPI. In six cases, a paper version was used in the whole school. This was at the request of the schools. The main reason was that the schools were extremely large and it would take too much time to have the pupils take the computer questionnaire individually, instead of using a class-administered paper questionnaire. In one case of these six, the whole school burned down and there were not enough computers available in the temporary building. In only one case, the school preferred PAPI because they felt they had too limited computer experience. In two

additional schools one class used a paper questionnaire due to a hardware problem in those classes, the other classes used the computer version.

Secondly, we compared the CSAQ and PAPI groups on background characteristics of the school and the class. The schools in the PAPI condition were, in general, larger than the schools in the CSAQ condition ($p = .05$). This is not surprising, as a large number of pupils was cited by schools as the main reason to request PAPI. However, the CSAQ and PAPI schools did **not** differ on other important variables, such as denomination, type of school, and urbanization. Nor did they differ regarding the socio-economic status of parents and age of pupils attending. Most important of all, according to the stated opinion of the teachers and other staff at the schools, the CSAQ and the PAPI schools do not differ in seriousness of bullying and victimization, which were the main topics of the pupils' questionnaire.

When we concentrate on the classes, we see that the teachers of CSAQ and PAPI classes do **not** differ in sex, education, teaching experience, job function (part-time versus full-time), or class level. There was a significant difference in the age of the teacher, the teachers in the PAPI condition being on average five years older ($p = .03$).

Although the two groups in this mode comparison were not randomly assigned, there were only minor differences between the CSAQ group and the PAPI group. We will come back to these differences when discussing the results.

3. Data Collection

3.1. Questionnaires

3.1.1. Survey on bullying

Two questionnaires were developed to measure the extent of and experiences with bullying in primary schools: a pupils' questionnaire and a teachers' questionnaire. The pupils' questionnaire focussed on attitudes regarding bullying, handling of reported bullying by teachers and parents, and actual bullying, either as victim or as active culprit. This questionnaire consisted of a total of 99 questions. The teachers' questionnaire consisted of a total of 130 and addressed the attitudes, perceptions and interventions of the teacher, regarding bullying in school or in the class. For the mode comparison (aim 2, data quality) we focus on the pupils' questionnaire, and the efficacy of computer-assisted self-administered questionnaires for children.

A Disk by Mail version of the questionnaires was developed using the CI3 program (Sawtooth 1994). Range checks were defined for all questions, and questions were randomized within blocks of related questions. A special code (9) was defined for ''do-not-know;'' this possibility did not appear on the screen, but was stated in a simple, separate instruction that was lying beside the computer. To accommodate this special population of young, primary school pupils, the possibility was created to temporarily stop when a child was tired or when the teacher needed a pupil, and resume at a more convenient time. A paper version of both questionnaires was available as back-up, and was used in a subsample of schools.

Also, a paper-and-pencil questionnaire was available for an additional sociometric study of the pupils. In the questionnaire the pupils were asked about their position in

the class, their social networks, and their favorite persons in the class. This sociometric test was part of a special study that was completely done in paper-and-pencil.

3.1.2. Evaluation of CSAQ

To investigate the feasibility of a computer-assisted data collection for pupils in primary schools (aim 1), we developed a short evaluation questionnaire for the teachers in the CSAQ condition. The first part of the questionnaire focussed on the acceptance of computer-assisted data collection of both pupils (how did they react) and teachers (what do you prefer). In the second part we asked about general problems of the pupils when completing the questionnaire (e.g., language problems, meanings of words used in questions), and special computer problems of pupils (e.g., keyboard use, readability of screen, understanding computer messages). The third part asked about logistical and technical problems (e.g., install the questionnaire on school computer) and background variables (e.g., computer experience of teacher).

To compare full costs for the two modes (aim 3), we used the financial records of the project manager.

3.2. Fieldwork and logistics

During the actual data collection we tried to reduce the burden for the teachers as far as possible, especially concerning the technical aspects of the data collection. We kept instructions simple and preprogrammed all main activities. In short: we aimed at a fool-proof system with accessible help in case of problems. This was implemented in the following way.

In March 1995 a small package, consisting of two or more disks (depending on the number of computers), three short printed instructions, and an accompanying letter, was sent to the teachers of the participating schools. The disk contained the interviewing program and some simple ''batch-files'' for starting the questionnaires, pausing and resuming, saving the data, and making back-ups. Two of the printed instructions were for the teacher: one gave instructions on how to start up the pupils' questionnaire, one gave instructions on how to start up the teachers' questionnaire. The third instruction, a yellow page with eight points in large letters, was developed for the pupils. This instruction was always lying beside the computer. Main points in the instruction were the use of <enter> and <back space>, and an explanation of the ''beep'' which was heard when a child gave an out of range answer or used <enter> without giving an answer. The instruction also stated that they were allowed to type in ''9'' if they REALLY could not give an answer to a specific question. Both teachers and children were reassured regarding the confidentiality of the individual answers.

A telephone help desk was operating, and people were on stand-by to go to a school with problems. Also several university laptops were available as back-up or as additional computers for very large schools. In eleven cases, one or two laptops were delivered to the school and assistance was offered when necessary. In one case, an assistant went to the school to give general support. This school had specifically asked for assistance because they were very worried as to whether they were capable of doing the ''computer things.'' The data collection was fully completed by August 1995.

Pupils were assigned to the computer by the teacher. Usually this was done at times when pupils had finished a learning task and were not needed by the teacher. Whenever a teacher needed a pupil, the interview program could be paused and then restarted at a more convenient moment.

It should be emphasized that all *eligible* pupils did complete the questionnaire, also the less literate ones. In very exceptional cases a pupil was not included. This concerned students who had recently immigrated to Holland and did not yet speak the language adequately. Using a paper-and-pencil questionnaire would not have changed the situation and these pupils also would not have been included in a standard study.

To evaluate the procedure (aim 1), all participating teachers were contacted after the data collection was finished. The teachers received a personalized report based on the results of their class, and at the same time they received a short paper-and-pencil questionnaire asking them to evaluate the data collection and the use of computer questionnaires. In this case we favored a standard paper-and-pencil questionnaire to avoid selective nonresponse in the group who did NOT favor Disk by Mail or had negative experiences. This mailing resulted in 128 completed questionnaires out of a total of 211 mailed out.

We sent out one complete reminder shortly before the Christmas holidays with a Christmas card personally signed by the project director. Although the last week of December is in general not a good period for mail surveys (cf. Dillman 1978), teachers indicated that this was the only period they had time. In fact this reminder worked extremely well and resulted in an additional 67 questionnaires. This brings the total response rate on the evaluation questionnaire to 195 questionnaires (92%).

4. Results: Acceptance, Data Quality, and Costs

4.1. Acceptance

The first aim of this study was to test the feasibility of computer-assisted data collection for pupils in primary schools, and especially the acceptance of this new method by both teachers and pupils. Criteria were the pupils' reactions and views as reported by the teachers, the teachers' own reactions, and the occurrence of technical problems. In a special evaluation, the teachers were asked about the experiences of their pupils, their own experiences, and the (computer) problems they encountered during the data collection. It was not feasible to do a full-scale evaluation including all 5,872 pupils; therefore the evaluation is based on teachers' reports and only gives a general impression of the acceptance by the pupils.

According to their teachers, in only a very small minority of the cases (0.5%) did the pupils not like CSAQ. A clear majority (68%) reported that not only did their pupils accept the method, but they also really enjoyed answering questionnaires by computer. When asked about the emotional burden of the questionnaire, teachers had the impression that asking these sensitive questions by computer was less stressful for their pupils with CSAQ (45%) or the same as with PAPI (52%).

The teachers also reported few problems during the data collection. In most cases (81%) they were asked for technical assistance *only once* during the entire data collection period.

We also asked the teachers how often they were called because the pupils had difficulty in answering the questions (e.g., did not understand the meaning of a word). Although we did our utmost to keep the questions simple, children had more problems with the questionnaire itself than with the computer: in 68% of the cases teachers were asked to explain the questions only once during the data collection period. The same picture emerges when we look at questions about problems with regard to the screen and using the keyboard. When we correct for reading problems, no clear problems were detected by the teachers. In sum: the use of a computer did not result in many additional problems for the pupils. Problems that did occur were mostly standard reading problems and problems connected with the understanding of question wording: the same problems one encounters using a pen-and-pencil questionnaire.

We also inspected correlations between answers on the evaluation questionnaire and background characteristics of the teachers. Contrary to our expectations we discovered that the older teachers, who reported that they did not have much computer experience, also reported slightly fewer problems during CSAQ, and were slightly more positive. Also, the lower grades with the young pupils had slightly more fun with the computer. On the other hand, the younger pupils did need more time to finish the questionnaire than the older pupils in the higher grades. Background variables such as school size, school denomination, teachers' sex, or teachers' education did not correlate with positive or negative reported experiences.

In sum: no clear pattern exists between background variables of school, teacher, or class and the results of the evaluation. There were two exceptions, less computer-experienced and older teachers evaluated CSAQ more positively, and younger pupils were reported as more enthusiastic. This is a very gratifying result, as these are precisely the groups that were expected to have the most problems with a computer assisted questionnaire. Consulting the computer-logs, we also found that the younger pupils took slightly longer to finish the questionnaire. We only have anecdotal evidence suggesting that also in the paper-and-pencil case younger children need some more time.

The results of the evaluation are stimulating for researchers who plan a computer survey at schools. The experiences are positive, older teachers react in a positive way and young children enjoy it. Very few problems were named by the teachers. Problems that were encountered were general language problems (e.g., meaning of a word), not computer problems. When implemented with care, DBM can be a success in primary schools.

4.2. Data quality: CSAQ versus PAPI

The second aim of this study was to determine if the use of CSAQ improves the quality of the resulting data when *children* are surveyed. Using a quasi-experimental design we compared the results of CSAQ with those of a paper-and-pencil questionnaire. Criteria for data quality were the amount of missing data, psychometric reliability of multi-item scales, and self-disclosure.

We first concentrated on the *missing values* in each condition. One of the most consistent findings in the literature is that computer-assisted data collection reduces the item nonresponse to both factual and opinion questions (Nicholls, Baker, and Martin 1997). This is mainly due to the elimination of omissions of applicable items by mistake.

We constructed a general indicator for item nonresponse, based on the percentage of missing values on the total of 111 numerical variables *all* pupils were presented with. Questions that only reached subgroups through skipping were excluded. The questions on which the missing value indicator was computed ranged from neutral background variables to more sensitive questions about bullying. Neither in the CSAQ nor in the PAPI condition was an explicit ''do not know'' or ''no answer'' option offered. It should be noted that in the CSAQ condition the children had the opportunity to hit a special key whenever they did not want to answer a question, and in the paper version they could simply skip the question.

A far higher percentage of missing values occurred in the PAPI condition ($p = .00$). In the CSAQ group the mean percentage of missing values was 5.7 (standard deviation = 3.4), while in the PAPI condition it was 14.1 (standard deviation = 25.0). These results suggest that not only the average amount of missing data is less in computer-assisted data collection, but also the individual differences, indicated by the standard deviation, are less. This could be attributed to the fact that with a paper questionnaire children who are not concentrating on the task or who are careless can easily skip a question or even a whole page by mistake. CSAQ forces children to be more precise.

In the literature on computer-assisted data collection an often cited disadvantage of such data collection is the ''itemization'' of the questionnaire: questions appear one by one on the screen instead of grouped in meaningful clusters on pages (cf. De Leeuw et al. 1995; Nicholls et al. 1997). This itemization could disturb the concentration of the children, resulting in less consistency in the computer-assisted version and so threatening the reliability of the results. In psychological testing explicit attention has been paid to the potential influence of computer administration on the data. The American Psychological Association demands that computerized versions of tests should be proven to be reliable and psychometrically equivalent before they may be used instead of the paper-and-pencil version. Also in survey research, psychometric reliability is considered an important aspect of data quality (Alwin and Krosnick 1991; Krosnick and Fabrigar 1997). Therefore, as second indicator for data quality, we used Cronbach's coefficient alpha (Cronbach 1951). This classical measure of *psychometric reliability* is an indicator of the consistency of the answers on a multi-item scale.

We computed Cronbach's alpha for two multi-item tests in the pupils' questionnaire, and compared these for the CSAQ and PAPI groups. For details of the statistical procedures, see De Leeuw (1992). The first test consisted of 21 questions about badgering and bullying (e.g., I sometimes would like to hit or kick other children; I sometimes would like to say nasty and mean things to other children in my class). The second test consisted of 20 questions about victimization (e.g., Other children often hit or kick me).

There was no statistically significant difference in reliability on the badgering/bullying test between the groups ($p = .12$). For the CSAQ condition the coefficient alpha was .88, and for the PAPI condition it was .90. There was a small statistically significant difference for the victimization test ($p = .02$). In the CSAQ condition the coefficient alpha was .90, and in the PAPI condition alpha was .87. Contrary to our expectations, we found no strong evidence that CSAQ lowers the consistency of the answers.

As third criterion for data quality we used self-disclosure on sensitive questions. Adults tend to give more open and less socially desirable answers when the data are collected

by means of a computer (De Leeuw et al. 1995; Nicholls et al. 1997; Weisband and Keisler 1996). We investigated if the same tendencies are found when children are respondents.

Regarding *openness* and *self-disclosure* we looked at the answers on both the badgering/bullying test and the victimization test. Children in the CSAQ condition reported that they were actively involved in more badgering and bullying than did children in the PAPI condition ($p = .00$). The mean score for the CSAQ condition was 30.5 (standard deviation = 9.6), while the mean score for the PAPI condition was 27.7 (standard deviation = 9.2). In the CSAQ condition also more victimization was reported ($p = .00$). The mean score on the victimization test was 26.4 (standard deviation = 9.8) for the CSAQ condition and 23.1 (standard deviation = 8.6) for the PAPI condition.

It could be argued that both groups really differ on victimization and bullying. However, this seems very unlikely, as there is nothing in the reports of the teachers and other staff of the schools to indicate that the CSAQ and the PAPI schools differed in seriousness of bullying and victimization. But we could also test this more stringently. From research on bullying and victimization it is known that it is the rejected pupils who are bullied, not the popular ones. From the same research it is known that bullies are not the most popular children and that they have a controversial status in the class (Perry, Kusel, and Perry 1988; Newcomb, Bukowski, and Pattee 1993; Hox and Van Hattum 1996). Therefore, if the children in the CSAQ condition were *really* more bullied and also if they bullied more than the children in the PAPI condition, the CSAQ children should also be less popular than the PAPI children. On the other hand, if the children in the CSAQ condition are only more open and honest about bullying than the PAPI children, the CSAQ children should be at least as popular as the PAPI children.

Fortunately, for all the pupils data were available on a sociometric test (Syracuse Amsterdam/Groningen Scale: SAGS; cf. Van Hattum 1997). This is an objective instrument to determine the social structure and networks in a group. Every group member gives a judgment of all the other individuals and will be judged by the others. Based on these judgments, popularity scores for each pupil can be computed. In all cases (i.e., for both the CSAQ and the PAPI groups) these data were collected in the classroom by means of a traditional paper-and-pencil test. This means that differences in sociometric scores cannot be attributed to a method effect. When we look at the CSAQ and PAPI groups, we see that the CSAQ-pupils are even *more popular* in their group than the PAPI pupils (CSAQ: mean = 2.3, standard deviation = .55; PAPI: mean = 2.2, standard deviation = .41; $p = .00$). Although the children in the CSAQ condition are on average somewhat more popular, they also report more victimization and bullying. This points to more openness and honesty with CSAQ. This is confirmed when we look into social desirability.

The main pupils' questionnaire also contained a short test measuring the tendency to give *socially desirable answers*, a high score on this nine-item-test indicating that a child has the tendency to give honest, socially undesirable answers (cf. Van Hattum 1997). There was a significant difference ($p = .00$) between the two conditions. Children in the CSAQ condition gave slightly more undesirable answers (mean = 30.6; standard deviation = 5.1) than children in the PAPI condition (mean = 29.9; standard deviation = 5.5).

In sum: the computer-assisted version of the children's questionnaire resulted in better

data quality, as indicated by fewer missing data and more openness and self-disclosure. Fortunately, the use of the computer did *not* lead to less consistency in the answers. Thus **no** evidence of the often feared itemization of the answers.

4.3. Cost

Besides data quality, cost is an important factor for consideration (cf. Groves 1989). Cost comparisons are always difficult. To give a fair comparison, we first present the costs we had, and compare these with the costs we should have had if we had done the same survey by paper and pen. Note that the costs of sampling, of developing the questionnaire, and of keeping account of the returned questionnaires are not taken into account; these would have been approximately the same in both cases.

The full extra costs for a first DBM survey were calculated, which gives us realistic figures. In fact, it was the first time that a DBM survey was done at the Educational Research Center of the University of Amsterdam. Of course, mastering the software and the actual programming depends on the experience of the researchers and the user-friendliness of the software. In this study, the actual programming in CI3 was done by a graduate student with only standard computer experience (e.g., word-processing, SPSS), who had to learn from scratch. Two of her supervisors, however, had extensive knowledge of programming and computer-assisted methods.

Costs are presented in Dutch guilders (f) with the equivalent in dollars at the current exchange rate in brackets.

Costs For Computer-Assisted Self-Administered Questionnaire (CSAQ)

	(f)	USD
CI3 program (Sawtooth)	4,000.–	(2,338.–)
Computer disks (500 * f1.95)	975.–	(569.89)[5]
Programming (including learning CI3, actual programming of the questionnaires, testing: 10 days * f420.30)	4,203.–	(2,456.65)
Help desk (3 days * f420.30)	1,260.90	(737.–)
Postage (including return mail 150 * f6.40)	960.–	(561.12)
Large envelopes (150 * f0.24)	36.–	(21.04)
Protective covers for disks (150 * f0.23)	34.50	(20.17)
Total DBM	**11,469.40**	**(6,703.87)**

Costs For Paper Mail Survey

	(f)	USD
Printing children questionnaire (6,800 * f2.72)	18,496.–	(10,810.91)
Printing teachers questionnaire (300 *f2.98)	894.–	(522.54)

[5] 467 disks were returned with completed questionnaires and can be reused. This lowers the costs when DBM surveys are done on a regular basis. Also, the costs of the program are one-time costs. In this comparison the **full** costs for disks and programming of a **first** DBM survey are counted.

Mailing[6]:		
Postage (305 * *f*9.– * 2)	5,490.–	(3,208.91)
Boxes (305 * *f*4.50)	1,372.50	(802.23)
Envelopes (305 * *f*0.24)	73.20	(42.79)
Data entry (6,628 returned questionnaires * *f*1.50)	9,942.–	(5,911.10)
Making system file (e.g., variable names, labels)	300.–	(175.35)
Total for PAPI	**36,567.70**	**(21,373.83)**

This brings the average cost in this survey for a completed questionnaire to *f*1.73 (1.01 US dollars) for a Disk by Mail survey and *f*5.52 (3.22 US dollars) for a paper-and-pencil mail survey.

5. Summary and Discussion

In this article we have presented the results of a Disk by Mail survey among children. Research on computer-assisted data collection has concentrated on adult subjects, and the main advantages (less item nonresponse, and slightly more self-disclosure) only generalize to an adult population. But the advantages of CSAQ are not limited to surveys of the adult population. We were able to show that:

(1) A Disk by Mail survey can be successfully implemented at Dutch primary schools.
(2) Children from the age of eight years on can successfully complete a computer-assisted self-interview, and enjoy it.
(3) Data quality in the computer-assisted group was better than in the paper-and-pencil group.
(4) DBM results in considerably less costs for each completed questionnaire than in the case of a PAPI mail survey.

It should be noted that due to practical considerations no full randomized experimental design could be used in the mode comparison; a quasi-experimental design was used instead (see Section 2.3). But we checked for potential differences between the groups and took these into account in the analyses. It seems very unlikely that the results are caused by differences between schools, and we are convinced that DBM can be used successfully to collect data from school-age populations. CSAQ can increase the data quality for surveys of children, as compared with paper-and-pencil self-administered surveys, without adding to the costs.

However, computer-assisted data collection is no panacea. To fully benefit from its advantages, it is necessary to do almost everything that is needed with a good paper-and-pencil questionnaire. In addition, one has to add extra effort in computer implementation and testing, in designing ergonomic screen layouts, and in targeting the CAI implementation to the specific population. Of utmost importance is a clear, short, paper guideline with instructions. When something does go wrong, help-functions often confuse

[6] This estimate is based on the cheapest solution, in which a package is sent by bulk mail to the school-classes. In this package an envelope will be included with instructions and stamps for the return mail. In our PAPI condition we in fact used a more expensive mailing out (no bulk rate).

unexperienced respondents, (see also De Leeuw and Nicholls 1996; De Leeuw, Hox, Kef, and Van Hattum 1997).

Some advantages of computer-assisted data collection are attributable to *technological* factors of well-designed and thoroughly tested computer questionnaires. For instance, it is not possible for a respondent to skip a question by mistake when using a computer. Our data clearly illustrate this: in the CSAQ condition the mean percentage of missing values was 5.7% (standard deviation = 3.4%), while in the PAPI condition it was 14.1% (standard deviation = 25.0%). It is interesting to note that not only the average amount of missing data is less, but also the individual differences, indicated by the standard deviation, are smaller when using computer-assisted data collection. This could be attributed to the fact that with a paper questionnaire children who are not concentrating on the task or who are careless can easily skip a question or even a whole page by mistake. CSAQ forces children to be more precise.

We also found that the younger children needed slightly more time to complete the computer questionnaire than the older children. Research concerning adults has shown that question texts are harder to read on monitor than on paper, and that ''speed'' tests and time constraints can threaten the data quality in CSAQ. (For an overview see De Leeuw et al. 1995.) In general it is wise to allow for ample time for the completion of a computerized questionnaire or interview. Regarding our findings, it is prudent to be especially careful in this when dealing with young children.

Other advantages can be attributed to *psychological* factors. Even when compared with paper-and-pencil self-administered questionnaires, the computer-assisted version resulted in more self-disclosure and less socially desirable answers. A computer seems to enhance the feeling of privacy and to underline the confidentiality of the answer (cf. Beckenbach 1995). Answers written down in a paper questionnaire are there for *everyone* to read, while answers that are typed in disappear from the screen and are in the perception of the respondent difficult to trace back by others. Of course much depends on the implementation of the survey. In our case, computers were standing in secluded corners and pupils went there individually. Also we took care that after an answer was typed in, it disappeared from the screen. There was no chance someone (e.g., a teacher, other pupils, assistants) could see what the respondent typed. When one is using computer-labs or other settings where students sit close to each other, the situation is different (cf. Beebe et al. 1998). When implementing a survey in these settings, one should take precautions that no-one can ''peep'' at the answers. The researcher should try to enhance the privacy of the students as far as possible, for instance by enlarging the distance between computers, by placing half-screens, or by effectively masking the answers that appear on the screen. Similar measures should be taken when using a CSAQ approach in a computer-assisted interview with children. The interviewer should explain the procedure carefully and then make it very clear (going to the other end of the room, looking out of the window, turning his/her chair) that he or she cannot see what is being typed in (see also De Leeuw et al. 1997 for some examples with blind adolescents).

In this study we also compared costs and were able to show that in our case the DBM survey cost about one third of what an equivalent PAPI mail survey would have cost. This is not always the case and depends on the ratio of fixed costs, which are independent of the size of the project (e.g., preparing and programming the questionnaire), and flexible costs,

which are a function of size (e.g., mailing costs, printing and editing of questionnaires)[7]. In small scale studies a paper-and-pencil survey may be more cost-efficient: the same observations have been made regarding CAPI and CATI studies (see for instance Bond 1991; Weeks 1992).

We can conclude that the often cited advantages of computer-assisted data collection also hold when the subjects are children. We should however point out that Dutch children are rather computer-literate and that the Ministry of Education stimulates the use of computers at Dutch primary schools. Furthermore, it should be noted that illiteracy was *no* problem in our population; the children could read and write on the basic level required for our questionnaire. Nevertheless, we did encounter some problems with the understanding of words used in the questionnaire. These were not problems caused by the computer or the screen: one would encounter the same problems using a paper-and-pencil questionnaire. The problems were misunderstandings of the meaning of words or not knowing what the actual meaning was. To put it more polemically, not computer technology, but the language that is used, may cause problems in interviewing children.

When surveying children in general, be it in oral interviews, with paper tests, or with computers, one should take extreme care in the construction and wording of the questionnaire and the instructions. When developing questionnaires or interview schedules for children, one should take into account the cognitive development and verbal capacities of the age-group (cf. Scott 1997; De Leeuw and Otter 1995). Pretests of questions are extremely important. One should always check whether children of a certain age-group understand the question the way it is intended by the researcher, for instance by applying ''cognitive laboratory techniques'' (see Forsyth and Lessler 1991 for a review of methods; Scott 1997 for an application with young children).

Also, it would be extremely interesting to develop new computer-assisted data collection and testing methods for young children with a still limited reading capacity. The recent developments in Audio-CASI (cf. O'Reilly, Hubbard, Lessler, Biemer, and Turner 1994) in combination with attractive graphics and simple mouse-oriented computer-human interaction, shows promise for future child-friendly surveys.

6. References

Alwin, D.F. and Krosnick, J. (1991). The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Characteristics. Sociological Methods and Research, 20, 139–181.

Aquilino, W. (1994). Interview Mode Effects in Surveys of Drug and Alcohol Use. Public Opinion Quarterly, 58, 210–240.

[7] One reviewer pointed out a hidden cost: the time of the teacher. In a paper-and-pencil mode, the questionnaire can be administered to a whole group. In this DBM application, the teacher had to allocate individual children to computers. We do not have time budget data available to formally test this. However, informal contacts with the schools suggest that the teachers worked rather efficiently. In the Dutch educational system, pupils work individually or in small groups a lot of the time. These periods were used to complete the computer questionnaire. After a pupil had finished a task correctly, he or she could work on the computer and complete the questionnaire. Furthermore, it should be noted that according to official regulations, large schools have more computers available than small schools, and could complete the DBM in the same time.

Beckenbach, A. (1995). Computer Assisted Questioning: The New Survey Methods in the Perception of the Respondent. BMS, 48, 82–100.

Beebe, T.J., Mc Rae, J.A., Anderson, R.E., and Fulkerson, J.A. (1998). An Evaluation of Computer-Assisted Self-Interviews in a School Setting. Public Opinion Research, 62, 623–632.

Besag, V.E. (1989). Bullies and Victims in Schools. Milton Keynes: Oxford University Press.

Bond, J. (1991). Increasing the Value of Computer Interviewing. Proceedings of the ESOMAR-conference.

Cook, T.D. and Campbell, D.T. (1979). Quasi-experimentation. Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Cronbach, L.J. (1951). Coefficient Alpha and the Internal Structure of Tests. Psychometrika, 16, 297–334.

De Leeuw, E.D., Hox, J., Kef, S., and Van Hattum, M. (1997). Overcoming the Problems of Special Interviews on Sensitive Topics: Computer Assisted Self-interviewing Tailored for Young Children and Adolescents. Proceedings of the 1997 Sawtooth Software Conference. Sequim, Wa: Sawtooth Software Inc.

De Leeuw, E.D. and Nicholls, W.L. II (1996). Technological Innovations in Data Collection: Acceptance, Data Quality, and Costs. Sociological Research Online, 1, 4, (http://www.socresonline.org.uk/socresonline/1/4/leeuw.html).

De Leeuw, E.D., Hox, J.J., and Snijkers, G. (1995). The Effect of Computer-assisted Interviewing on Data Quality: A Review. Journal of the Market Research Society, 37, 4, 325–344.

De Leeuw, E.D. and Otter, M.E. (1995). The Reliability of Children's Responses to Questionnaire Items: Question Effects in Children's Questionnaire Data. In Advances in Family Research, eds. J.J. Hox, B.F., van der Meulen, J.M.A.M. Janssens, J.J.F. ter Laak, and L.W.C. Tavecchio. Amsterdam: Thesis Publishers.

De Leeuw, E.D. (1992). Data Quality in Mail, Telephone and Face-to-Face Surveys (chap. 3). Amsterdam: TT-publikaties.

Dillman, D.A. (1978). Mail and Telephone Surveys; The Total Design Method. New York: Wiley.

Forsyth, B.H. and Lessler, J.T. (1991). Cognitive Laboratory Methods: A Taxonomy. In Measurement Errors in Surveys, eds. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman. New York: Wiley.

Groves, R.M. (1989). Survey Errors and Survey Costs. New York: Wiley.

Heenan, S.A. (1992). The Problems of Carrying out Small Scale Research into Bullying in Schools. Unpublished Master thesis, University of East Anglia, School of Education.

Hochstim, J.R. (1967). A Critical Comparison of Three Strategies of Collecting Data from Households. Journal of the American Statistical Association, 62, 976–989.

Hox, J.J. and Van Hattum, M.J.C. (1996). De invloed van sociometrische positie op pesten in de klas. Modellen en analyseprocedures. (The Influence of Sociometrical Score on Bullying in the Class.) Amsterdam: University of Amsterdam, SCO-Kohnstamminstituut. (In Dutch.)

Krosnick, J.A. and Fabrigar, L.R. (1997). Designing Rating Scales for Effective Measurement in Surveys. In Survey Measurement and Process Quality, eds. L. Lyberg,

P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley.

Newcomb, A.F., Bukowski, W.M., and Pattee, L. (1993). Children's Peer Relations: A Meta-Analytic Review of Popular, Rejected, Neglected, Controversial, and Average-Sociometric Status. Psychological Bulletin, 113, 99–128.

Nicholls, W.L. II, Baker, R.P., and Martin, J. (1997). The Effect of New Data Collection Technologies on Survey Data Quality. In Survey Measurement and Process Quality, eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley.

O'Reilly, J.M., Hubbard, M.L., Lessler, J.T., Biemer, P.P., and Turner, C.F. (1994). Audio and Video Computer Assisted Self-Interviewing: Preliminary Tests of New Technologies for Data Collection. Journal of Official Statistics, 10, 197–214.

Perry, D.G., Kusel, S.J. and Perry, L.C. (1988). Victims of Peer Aggression. Developmental Psychology, 24, 807–814.

Saltzman, A. (1993). Improving Response Rates in Disk-by-Mail Surveys. Marketing Research, 5, 32–39.

Sawtooth, (1994). CI3 User Manual (edited by Harla Hutchinson and Margo Metegrano). Evanstone: Sawtooth Software.

Schwarz, N., Strack, F., Hippler, H-J., and Bishop, G. (1991). The Impact of Administration Mode on Response Effects in Survey Measurement. Applied Cognitive Psychology, 5, 193–212.

Scott, J. (1997). Children as Respondents: Methods for Improving Data Quality. In Survey Measurement and Process Quality, eds. L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin. New York: Wiley.

Siemiatycki, J. (1979). A Comparison of Mail, Telephone, and Home Interview Strategies for Household Health Surveys. American Journal of Public Health, 69, 238–245.

Sudman, S. and Bradburn, N.M. (1974). Response Effects in Surveys: A Review and Synthesis. Chicago: Aldine.

Tourangeau, R. and Smith, T.W. (1996). Asking Sensitive Questions; The Impact of Data Collection, Question Format, and Question Context. Public Opinion Quarterly, 60, 275–304.

Turner, C.F., Lessler, J.T., and Devore, J.W. (1992). Effects of Mode of Administration and Wording on Reporting of Drug Use. In Survey Measurement of Drug Use: Methodological Studies, eds. C.F. Turner et al. Washington DC: Government Printing Office.

Van der Ree, B. (1994). The COMENIUS Project for Primary and Special Education. Zoetermeer: The Netherlands Ministry of Education, Culture, and Sciences.

Van Hattum, M.J.C. (1997). Pesten. Een onderzoek naar beleving, visie en handelen van leraren en leerlingen [Bullying: An Empirical Study Among Teachers and Pupils of Primary Schools]. Amsterdam: University of Amsterdam. (In Dutch.)

Weeks, M.F. (1992). Computer-assisted Survey Information Collection: A Review of CASIC Methods and Their Implication for Survey Operations. Journal of Official Statistics, 8, 445–466.

Weisband, S. and Kiesler, S. (1996). Self Disclosure on Computer Forms: Meta Analysis and Implications. CHI '96, (http://www.al.arizona.edu/~weisband/chi/chi96.html).

Witt, K.J. and Bernstein, S. (1992). Best Practices in Disk-by-Mail Surveys. Proceedings of the 1992 Sawtooth conference, Sawtooth software: Evanston, Illinois, U.S.A.

Zandan, P. and Frost, L. (1989). Customer Satisfaction Research Using Disks-by-Mail. Proceedings of the 1989 Sawtooth conference, Sawtooth software: Evanston, Illinois, U.S.A.

Received February 1997
Revised March 1998