Journal of Official Statistics, Vol. 17, No. 2, 2001, pp. 321–335

Item Nonresponse in Questionnaire Research With Children

Natacha Borgers [1] *and Joop Hox* [2]

This study investigates the effect of item and person characteristics on item nonresponse, for written questionnaires used with school children. Secondary analyses were done on questionnaire data collected in five distinct studies. To analyze the data, logistic multilevel analysis was used with the items at the lowest and the children at the highest level. Item nonresponse turns out to be relatively rare. Item nonresponse can be predicted by some of the item and person characteristics in our study. However, the predicted response differences are small. There are interactions between item and person characteristics, especially with the number of years of education, which is used as a proxy indicator for cognitive skill. Young children do not perform as well as children with more years of education, by producing more item nonresponse, but their performance is still acceptable.

Key words: Partial nonresponse; special population; item characteristics; respondent characteristics.

1. Introduction

Increasingly, survey researchers who are interested in perspectives, attitudes, and behavior of children, collect this information directly from the children themselves, and not by proxy reporting other sources of information (Scott 1997). However, methodological knowledge on how to best survey children is still scarce. Our study investigates the quality of questionnaire data obtained from children, especially with respect to partial nonresponse (item nonresponse).

We use Krosnick's satisficing theory (1991), which posits three factors that affect the process of answering questions. The first is the motivation of the respondent to perform the task, the second is the difficulty of the task, and the last is the respondent's cognitive ability to perform the task. The satisficing theory elaborates on a standard question-answering process-model developed by Tourangeau and Rasinski (1988) (cf. Cannell, Miller, and Oksenberg 1981; Krosnick, Narayan, and Smith 1996; Sudman, Bradburn, and Schwarz 1996; Schwarz, Knäuper, and Park 1998). It identifies two processes that explain differences in response quality, namely optimizing and satisficing. Optimizing means that the respondent goes through all the cognitive steps needed to answer a survey question. Satisficing means that a respondent gives responses that appear reasonable or acceptable, without going through all the steps involved in the question-answering process. Satisficing is related to the motivation of the respondent, the difficulties of the

[1] Faculty of Social and Behavioral Sciences, University of Amsterdam, Wibautstraat 4, NL-1091 CM Amsterdam, The Netherlands. E-mail: nborgers@educ.uva.nl
[2] Utrecht University, The Netherlands.

task and the cognitive abilities of the respondent. Difficult questions and/or low cognitive abilities may lead respondents to provide a satisfactory response instead of an optimal one. Using a satisficing strategy results in less reliable responses than using an optimizing strategy.

The results of empirical research support the idea that question characteristics and respondent characteristics affect response quality (e.g., Krosnick 1991). Less research has been done on the effects of respondent characteristics on response quality. Many respondent characteristics (like motivation) are difficult to measure within a survey, and most of the attention is devoted to observable traits of the respondent, such as age and education (cf. Krosnick and Alwin 1987; Alwin and Krosnick 1991; Rodgers, Andrews, and Herzog 1989; Andrews and Herzog 1986, 1988). Results from different studies show that reduction in cognitive function is negatively related to response quality (Knäuper, Belli, and Hill 1997; Alwin and Krosnick 1991; Krosnick 1991).

Directly following from the satisficing theory is the existence of an interaction effect between respondent characteristics and question characteristics, which can be described as follows: the less cognitively sophisticated respondents are more sensitive to difficult or cognitively demanding questions, and their responses will be less reliable. There is some research directed at this interaction, which supports this hypothesis (Knäuper, Belli, and Hill 1997; Schwarz, Knäuper, and Park 1998).

The satisficing theory and the empirical results concentrate on adult respondents. Children are interesting subjects in this respect, because they are still developing their cognitive skills, and therefore cognitive ability varies considerably across children.

Most studies have used different measures of reliability as indicators of response quality. In our study, we investigated the effects on item nonresponse, which also depends on characteristics both of the respondent (De Leeuw 1999) and of the questions (Krosnick and Fabrigar 1997; Leigh and Martin 1987). Thus, our research question is: What are the effects of respondent characteristics and item characteristics (and their interaction effects) on item nonresponse produced by children in self-administered questionnaire research?

2. Method

2.1. Data sets

We use secondary analysis on five different data sets, collected in the field of educational research. These data sets consist of the responses on self-administered paper and pencil questionnaires, administered in class settings in the course of different educational research projects in the Netherlands and Belgium. There were no time limits for the children to answer the questions. The data sets are briefly described in the Appendix. We have data collected with ten questionnaires, which contain 37 different scales and 348 items, which were answered by 3,492 children. All questions are part of (sub) scales, lists of items that have the same format. Our sample is between 8 and 18 years of age ($\bar{X} = 12$, s.d. $= 2.2$). For all the children the number of years of education was known, the variation being from 4 to 14 ($\bar{X} = 8$, s.d. $= 2.2$). The total data set covered 1,322 boys and 1,349 girls, and 821 children for whom the sex is not recorded in the data file.

2.2. Coding scheme

Based on a literature study, we developed a computerized coding scheme (Borgers 1997) to code all items. A summary of the item characteristics and their operationalization is given in Table 1 (the full coding scheme is available from the first author).

All 26 codes measure an aspect of the extent to which an item is cognitively demanding. Some of these characteristics, e.g., item ambiguity, are to some degree subjective. By combining the results of multiple coders, the reliability of the composite rating can still be satisfactory (Stock 1994). In our case, two coders were used for such characteristics. For each characteristic the intercoder reliability (Cohen's kappa) was determined. If this was lower than 0.70, a third coder was assigned, and the final code is the mean rating of all coders. The mean intercoder reliability (Orwin 1994) was above 0.80, except for the four indicators for subjective question threat. However, these four indicators are combined into one rating. The reliability of the combined ratings[3] is used (Guilford 1954). This results in an intercoder reliability of .32, which is not very high, but sufficient for our purpose. A summary of all intercoder reliabilities and some descriptive statistics of the item characteristics are given in the Appendix.

2.3. Analysis

Our data can be viewed as hierarchical. In total there are four levels: items which are nested in scales, scales which are nested in questionnaires, and questionnaires which are nested within respondents. To analyze data with such a hierarchical structure, multilevel analysis is appropriate. Given the relatively small sample sizes at both the study level ($N = 5$) and the questionnaire level ($N = 10$), we use a two-level model to incorporate this nesting structure. In this model, items are the lowest and respondents the highest level.

The dependent variable in this study, item nonresponse, is measured as whether or not an item is answered by a child. This is a dichotomous variable: 0 = no answer, 1 = answer to the item. The item nonresponse variable is extremely skewed; most children answered most items. Such data violate several assumptions of the normal regression method, and the appropriate analysis model is logistic regression. Consequently, the analysis used here is a multilevel logistic regression model. For an introduction to multilevel modeling, see Bryk and Raudenbush (1992) and Hox (1995); the multilevel logistic model is described in Goldstein (1995).

The hierarchical structure of the data implies that the dependent variable is measured for every combination of a child, study, questionnaire, scale, and item. This produces an enormous data set, which is too large for the multilevel program Mlwin (Goldstein et al. 1998). Therefore, we made a preselection of the explanatory variables. Although all question characteristics were initially selected on a theoretical or empirical basis (Borgers 1999), this does not indicate the importance of expected effects. Therefore, we performed an empirical selection; for every explanatory variable one separate multilevel logistic regression analysis was carried out to determine if it had a significant contribution.

$$\alpha_{sum} = 1 - \frac{\sum \sigma_j^2 - \sum \sigma_j^2 \alpha_j}{\sum \sigma_j^2 + 2 \sum_{k>j} \sigma_j^2 \sigma_k^2 r_{jk}}$$

Table 1. Summary of the item characteristics, and operationalization

Comprehension and interpretation of the question	Operationalization
Question length	number of words number of sentences
Length of the introductory text	number of words/100
Readability	comprehensive readability index (high score = easy to read) technical readability index (high score = easy to read)
Ambiguity of the item	ambiguous versus unambiguous (if meaning of the question is not the same for all respondents)
Ambiguity of the response scale	ambiguous versus unambiguous (if meaning of the response scale is not the same for all respondents)
Double-barreled	double versus single
Complex constructions	complex versus simple (if item contains complex constructions like subordinate clauses, semicolons, etc.)
Negatively formulated items	negatively versus positively formulated (double negations)
Kind of information being asked	attitudes experiences opinions behavior attributions capacities
Retrieving relevant information from memory	
Complexity of the item	complex versus simple (information that is being asked for cannot directly be retrieved from memory)
Reference period	reference period versus no reference period
Numerical quantity	numeric versus not numeric response
Judging the retrieved information	
Subjective question threat, sensitivity	sum of 4 indicators[4]
Balance of the item	balanced versus unbalanced number of response options[5]
Position in the questionnaire	1st versus 2nd versus 3rd position in the questionnaire[5]
Communicate the final response	
Number of response categories	2, 3, 4, 5, 7 and 10 categories[6]
Offering midpoints	midpoint versus no midpoint offered
Offering Don't know filter	don't know filter versus no don't know filter
Scale labels	labeled versus partly labeled response scale

[4]1: Too personal for the respondent; 2: Too threatening for the respondent; 3: Rather not answer the question; 4: Hard to give an honest answer to the question.

[5]Every questionnaire was divided in three parts with an equal amount of questions, due to the different number of questions in the questionnaires.

[6]The number of response options is included as dummy variables, to leave open the possibility of nonlinear effects.

Nonsignificant contributions ($\alpha = 0.05$) were excluded from the final analysis. The only exceptions were three categorical variables, which are represented by dummy variables: position in the questionnaire, number of response categories, and the kind of information that is being asked for, respectively. For position in the questionnaire, two dummy variables were used: one for the second and one for the third position in the questionnaire. The second categorical variable in our study is the number of response options. Two response options is the base against which the remaining five options are compared. The last variable concerns the type of information that is being asked for. For this variable, we computed seven dummy variables with attitude items as the baseline. These categorical variables were all included in the model if at least one of the dummy variables made a significant contribution. However, the set of dummy variables that indicate type of information caused convergence problems in the preselection. Six out of the seven dummy variables did not differ significantly from the baseline, and we decided to exclude all except the significant dummy variable that indicates knowledge items. Missing observations for sex and age were replaced by the mean score, based on the known years of education within each questionnaire.

After this preselection, only the significant contributions were included in a combined model. Initially, all coefficients were treated as fixed. The following step was to model these coefficients as varying across children. We expect different effects of the question characteristics across children, and varying slopes for item effects indicating interaction effects with some child characteristic. Thus, the last step was introducing interaction variables to model the significant random coefficients. To test our hypothesis we need to include cross level interaction variables; the item characteristic interacting with the available child characteristics: years of education, age and sex. Our final model only contains the significant contributions.

3. Results

The results of our multilevel logistic regression analysis are presented in Table 2. The first column presents the results of the empty or intercept only model (cf. Hox 1995). In this column, we see that 86 per cent of the variance in the proportion of answered items is on the child level, and 14 per cent on the item level. This means that, in general, child characteristics are probably more important than question characteristics in predicting item nonresponse. The second column presents the logit transformed regression coefficients for our final model. All significant ($\alpha = 0.05$) contributions are shown. In the last column the expected proportions of responses are presented.

Two things are important when interpreting the results. First, the regression parameters are interpreted in terms of the underlying variate defined by the logit transformation. To understand the implications of the regression coefficients for the proportions we have to transform their values back to the original scale. The expected proportions of responses in Table 2 are based on the sum of the intercept and the separate regression coefficients. Thus, the numbers in the last column indicate the contribution of a specific item characteristic, compared to the intercept value of 0.945, with all other variables having the value zero. In all figures that follow, the effect of a variable is shown, with all other variables set to their overall mean value. This provides a more realistic picture.

Table 2. Logit transformed regression coefficients and expected proportion response for child and item characteristics on response (N = 197,552)

	Intercept only model	Final model	Expected proportion response
Fixed part			
Intercept	4.669 (.00)	2.843 (.00)	.945
Child characteristics			
Years of education		.465 (.00)	.964
Item characteristics			
Length of introductory text in words		.003 (.00)	.959 (per 100 words)
Ambiguity of response options		1.300 (.02)	.984
Knowledge		1.083 (.01)	.981
Social sensitivity of the item		1.490 (.11)[a]	.987
Position in the questionnaire			
Second position		−.311 (.00)	.927
Third position		−.596 (.00)	.905
# of response categories			
Three		−.811 (.00)	.884
Four		−.435 (.04)	.918
Five		−1.146 (.00)	.846
Seven		−1.819 (.00)	.736
Ten		−2.232 (.00)	.649
Interaction terms			
Years of education[a] ambiguity		−.336 (.00)	.925
Years of education[a] social sensitivity		−.242 (.04)	.928
Random part			
Variance at child level	21.060 (.00)	22.171 (.00)	
Ambiguity		14.011 (.00)	
Second position		30.725 (.00)	
Third position		37.318 (.00)	
Social sensitivity		63.927 (.00)	
Variance at item level	3.29		

Note. Values in parentheses are p-values.
[a]This coefficient is not significant at $\alpha = 0.05$ but is included in the model because it is part of a significant interaction.

Secondly, the regression coefficients of the item variables "ambiguity" and "social sensitivity" and the child variable "years of education" that are part of an interaction cannot be interpreted separately. In the following, we first discuss the results for the fixed regression coefficients, for those variables that are not a part of an interaction. Subsequently we will discuss the results of the variances of the regression coefficients across children, and the interaction effects.

3.1. Regression coefficients

The length of the introductory text has a positive effect on the response, as do knowledge items compared to all other types of information being asked for. Position in the questionnaire and number of response options have negative effects on responses. If we consider the expected response differences, increasing the length of the introductory text by 100 words raises the expected response by 1.5 per cent. For knowledge items as opposed to other items, this increase is 3.6 per cent. If the item is located in the second part of the questionnaire, the expected response decreases by 1.8 per cent compared to items in the first part of the questionnaire. The response on items in the third part will decrease by 4.0 per cent compared to items located in the first part of the questionnaire. In Figure 1 the results are visualized, transformed into expected response proportions.

The results indicate that there is a decrease in responses when an item is located in the second part, and even more when an item is located in the third part. However, the expected differences are very small.

In general, the more response options offered, the more nonresponse would occur. The results are presented in Figure 2.

The results indicate a decrease of responses with increased number of response options. However, the differences are again very small.

3.2. Variances and interaction effects

The lowest level (item) variance is not included in the second column of Table 1, because in logistic models this is always fixed to the variance of a logistic distribution with scale parameter equal to one. So only the child level variance can be interpreted.

There is significant variation across children for the effects of ambiguity of the item, position in the questionnaire and social sensitivity of the item. To test our interaction hypothesis we included cross-level interaction variables for position in the questionnaire,

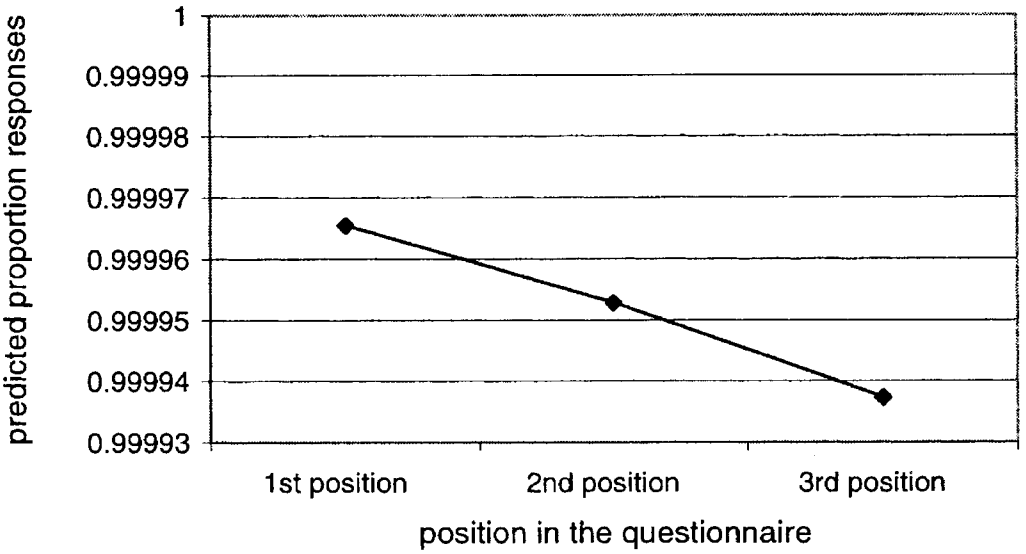

Fig. 1. Expected proportion responses for different positions in the questionnaire with all other explanatory variables equal to their overall mean

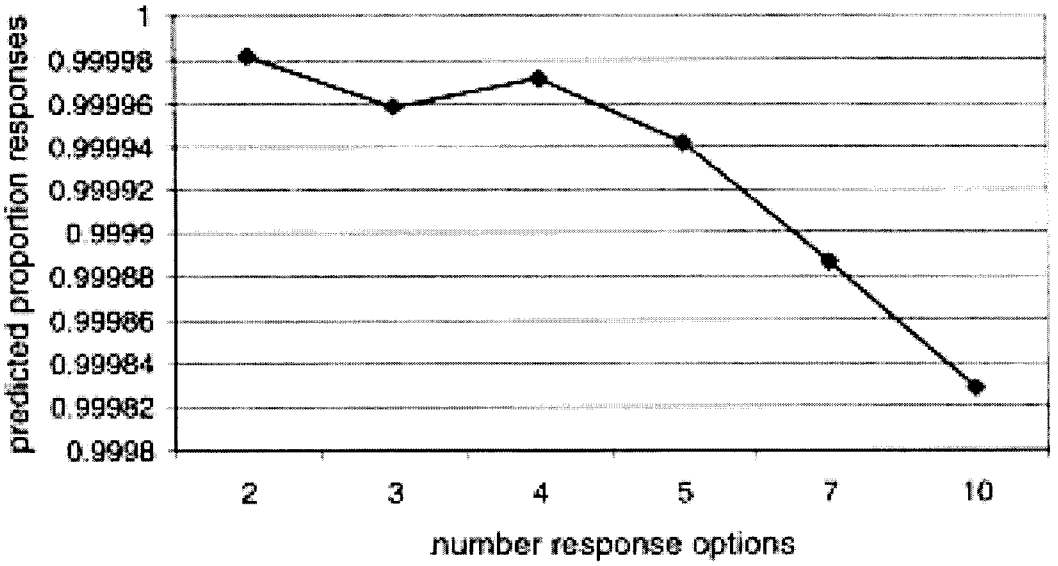

Fig. 2. Predicted proportion responses for different numbers of response options with all other explanatory variables equal to their overall mean

ambiguity of the response scale, social sensitivity and the available child characteristics age, sex and years of education. The interaction variables for position in the questionnaire and years of education, age and sex are not significant. The interaction effects between years of education and ambiguity of the response scale and between years of education and social sensitivity are significant.

The regression coefficients of the variables that are part of the interaction cannot be interpreted in isolation, and there is also the issue of centering (cf. Aiken and West 1991). Therefore, the predicted values back to proportions, and present the interaction effects in Figure 3 and Figure 4.

Items with an ambiguous response scale produce more response than those with unambiguous response scales. Interestingly, this effect does not occur with the youngest children. The youngest children do produce the most nonresponse but there is no difference between ambiguous and unambiguous response scales. After six years of education the

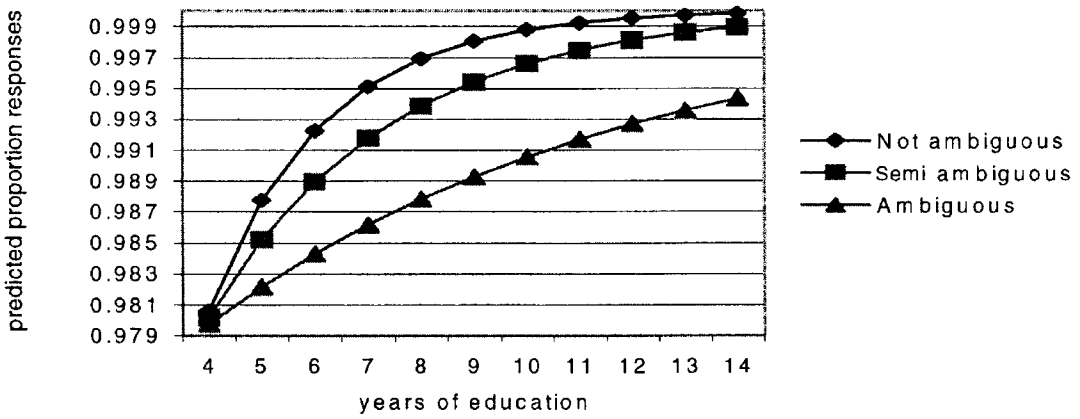

Fig. 3. Proportion answered items for different years of education and ambiguity of response scale with all other explanatory variables equal to their overall mean

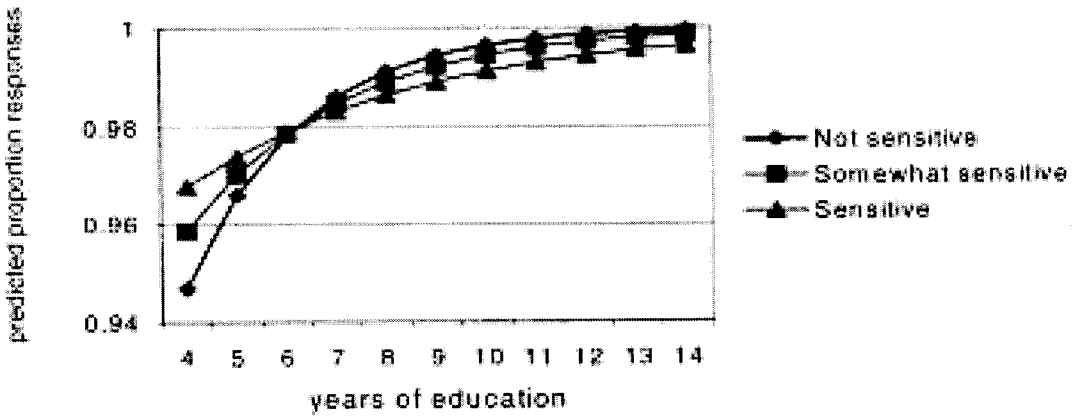

Fig. 4. Proportion answered items for different years of education and social sensitivity of items with all other explanatory variables equal to their overall mean

difference in proportion responses between ambiguous and unambiguous response scales appears. The difference becomes smaller after 13 years of education but is still there.

Only the youngest children, with up to six years of education, show differences in the effect of sensitive items on item nonresponse. This shows almost the same pattern as the effects of ambiguous response scales. Young children produce less item nonresponse with sensitive items compared to items that can be characterized as not sensitive. From the sixth year of education and up there is almost no difference.

4. Discussion

In general, it is difficult to predict item nonresponse by child characteristics, by item characteristics, and their interaction. However, even if we can model only little variation in nonresponse, we can still use the results to improve items with respect to item nonresponse. One can argue that increasing responses by one per cent is not worth the effort. However, since most statistical programs use listwise deletion, even a one per cent item nonresponse can result in a sizeable loss of information in the whole data set. For example: assume a regression analysis with one predictor, consisting of the sum score of ten items. If half of the respondents skip just one of the items, the overall proportion of response is an acceptable 0.95, but in the end the regression analysis will be based on only half the data set. That makes the effort of adapting items worthwhile, even if the proportion of responses per item increases only slightly.

Our study indicates some item characteristics that we may consider when designing a questionnaire for children. The first is the position of the item in the questionnaire. We cannot avoid item nonresponse by changing the position in the questionnaire but we can randomize item nonresponse by randomizing the position of items in a questionnaire. In that way the proportion of item nonresponse is randomized over items and not systematically the largest for the items in the third part of the questionnaire. The second is using a clear and extensive introductory text in a questionnaire, which prevents item nonresponse to some extent.

Adapting the number of response options can also prevent loss of information. Increasing the number of response options up to ten decreases the predicted proportion of

response, although the difference is small. This result is opposed to research on the effects of response options on the reliability of responses. The more response options offered, the higher the reliability of responses, with an optimum of around seven response options (e.g., Alwin 1997; Alwin and Krosnick 1991; Miller 1956). Since the importance of a higher reliability far outweighs the small loss in item response, we recommend the use of at least four or five response options.

Furthermore, one should avoid offering ambiguous response scales and sensitive items. In our study, the youngest children (with four and five years of education) produce less item nonresponse with sensitive items. In addition, these children do not produce more item nonresponse for items with ambiguous response scales, while children with more years of education do. These results are interesting, and are likely to be associated with the developmental stage of these children. Children in this age group are very literal in their interpretation of words (Holoday and Turner-Henson 1989). They are also limited in their language development, which implies limitations in comprehension and in verbal memory. It is likely that these children do not recognize the ambiguity of the words in the response scale. However, with older children, using ambiguous items does lead to an increase in the proportion of responses, and, more importantly, a decrease in the reliability of responses (De Leeuw and Otter 1995).

Our theoretical basis, Krosnick's satisficing theory, refers in particular to the effects of interactions between child and item characteristics on reliability of responses. In our research we tried to apply this theory to item nonresponse. In this line of reasoning, item nonresponse is an extreme expression of satisficing. Although item nonresponse is a rare phenomenon, and hard to predict, we found some interesting results. Some of these, such as the effects of the position of items in the questionnaire, the ambiguity of the response scale and the length of the introductory text, are in concordance with the satisficing theory. However, other results, such as the effect of negatively formulated items and double-barreled items, are not. Hence, we are not able to unequivocally confirm or reject the satisficing theory for item nonresponse. Skipping an item in a questionnaire may not be an expression of satisficing and may have its own mechanism. On the other hand, difficulties with predictions can be caused by the fact that skipping an item is a rare phenomenon. It is less difficult to explain the reliability of responses by child characteristics and item characteristics than it is in the case of item nonresponse.

Finally, we note that our results are based on a secondary analysis of available data. The effects appear plausible, but a study in which these variables are manipulated in an experimental design would allow stronger conclusions. In addition, all data sets are based on items that form a scale, collected using self-administered paper-and-pen questionnaires. Small mode differences in item nonresponse do exist (De Leeuw 1992), so our results may not apply to orally administered questionnaires. The fact that all our items are embedded in a scale, with similar items using the same response alternatives, may have contributed to the overall high level of response. This form of presentation induces a cognitive set in the respondents, which is beneficial for the quality of the responses. A series of independent question is expected to have higher levels of item nonresponse, and may be more sensitive to adverse item characteristics. Besides, all questionnaires were administered in a school setting. These settings could incite children to adopt a school test attitude. As a result children might

be less likely to satisfice and produce less item nonresponse than a setting outside the classroom.

The size of our data set made it necessary to make a preselection of the variables. Since this preselection is based on the significance level in a preliminary test, there is a risk of increased type I error. If we adjust our significance level to a more stringent ($\alpha = .01$) level, most of the effects in Table 2 are still significant. However, the interaction between years of education and social sensitivity is now clearly no longer significant. This means that we are no longer able to model the child level variance of the social sensitivity regression slope.

Appendix A

Summary of the data sets: Instrument; Scales (number of items)
(a full description is available from the first author).

Reading pleasure (Schoonen et al. 1993) 1 scale, 20 questions, 443 pupils

- Reading attitude (20)

Perceived competence (Van den Bergh 1995) 6 scales, 36 questions, 758 pupils

- Scholastic competence (6)
- Social acceptance (6)
- Athletic competence (6)
- Physical appearance (6)
- Behavioral conduct (6)
- Global self-worth (6)

Tutor program reading education/ZOS questionnaire (Fukkink 1996; Fukkink and Vaessen 1996)

4 scales, 66 questions, 800 pupils

- Reading attitude (27)
- (Self-)confidence with regard to explaining schoolwork to other pupils (13)
- Contact with fellow pupils (15)
- Sensitivity (11)

Environmental behaviour questionnaire (Lagerweij 1995)

8 scales, 75 questions, 699 pupils

- Environmental behavior (30)
- Communication environmental behavior at home (7)
- Environmental consciousness parents (5)
- Attitude environmental behavior (8)
- Social responsibility (10)
- Social undesirability (6)
- Environmental behavior at home (4)
- Knowledge environmental behavior (5)

Students' perspectives and efforts regarding school questionnaires (Peetsma 1991)

18 scales, 147 questions, 606 pupils

- Short term perspective regarding school and vocational career (6)
- Long term perspective regarding school and vocational career (6)
- Short term perspective regarding social relations (6)
- Long term perspective regarding social relations (6)
- Short term perspective regarding personal development (6)
- Long term perspective regarding personal development (6)
- Short term perspective regarding recreation (6)
- Long term perspective regarding recreation (6)
- Effort regarding English in the class (14)
- Effort regarding English homework (14)
- Effort regarding school in general (14)
- Effort regarding homework in general (14)
- Achievement motivation scale: negative fear of failure (15)
- Locus of control regarding personal effectiveness in achievement (6)
- Locus of control regarding social relations (5)
- Estimation of school instrumentality concerning vocational and school career in the future (5)
- Primary education in retrospect (6)
- Memories about the situation as a pupil in primary education (6)

Appendix B

Summary of the intercoder reliability for all questionnaires after corrections
Number of items = 348, Descriptive statistics

	Cohen's Kappa	mean (s.d.)	min.–max.
Number of words in the introductory text (/100)	1.00	1.69 (1.54)	0–3.9
Instruction to the question	1.00	All questions were coded 0; this characteristic is not included in the analyses	
Ambiguity of the question	.86	.36 (.47)	0–1
Ambiguity of the response scale	.90	.40 (.46)	0–1
Double-barreled question	.91	.15 (.35)	0–1
Complex constructions	.95	.21 (.40)	0–1
Negative formulated question	.91	.07 (.25)	0–1
Kind of information that is being asked for	1.00		1–9
Complexity of the question	.87	.54 (.49)	0–1
Reference period	1.00	.11 (.33)	0–2
Numeric quantity	99% corresponding codes	.00 (.04)	0–1
Too personal for the respondent Too threatening for the respondent Rather not answer the question Hard to give an honest answer to the question	.32	.07 (.16)	0–1
Position in the questionnaire	1.00		1–3
Balanced question	1.00	.41 (.49)	0–1
Number of answer categories	1.00	4.85 (1.6)	2–7
Neutral midpoint	1.00	.14 (.35)	0–1
Labeled response scale	1.00	1.73 (.44)	1–2
Don't know filter	1.00	.07 (.25)	0–1
Something else category	1.00	All questions were coded 0; this characteristic is not included in the analyses	

5. References

Aiken, L.S. and West, S.G. (1991). Multiple Regression: Testing and Interpretating Interaction. Newbury Park, California: Sage.

Alwin, D.F. (1997). Feeling Thermometers Versus 7-point Scales. Which are Better? Sociological Methods and Research, 25, 3, 318–340.

Alwin, D.F. and Krosnick, J.A. (1991). The Reliability of Survey Attitude Measurement. The Influence of Question and Respondent Attributes. Sociological Methods and Research, 20, 1, 139–181.

Andrews, F.M. and Herzog, A.R. (1986). The Quality of Survey Data as Related to Age of Respondents. Journal of the American Statistical Association, 81, 394, 403–410.

Bergh, B. van den (1995). Onderzoek "De leefsituatie van Kinderen op schoolleeftijd." Beknopte beschrijving van het onderzoek + vragenlijsten + bijlagen. Internal report, Brussel: Centrum voor Bevolkings- en Gezinsstudie.

Borgers, N. (1997). Codeerschema. Computer Program.

Borgers, N. (1999). Coding Scheme: A Technical Report. Unpublished report, Amsterdam, University of Amsterdam.

Byrk, A.S. and Raudenbush, S.W. (1992). Hierarchical Linear Model: Applications and Data Analysis Methods. Newbury Park: Sage.

Cannel, Miller, and Oksenberg (1981). Research on Interviewing Techniques. Field Experiment in Health Reporting, 1971–1977.

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

De Leeuw, E.D. (1999). Item Nonresponse: Prevention is Better than Cure. Survey Methods Newsletter, 19, (draft extract).

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, L.W.C. Tavecchio, Amsterdam: Thesis Publishers.

Fukkink, R. and Vaessen, X. (1996). Stap Door! Verslag van het experimentele jaar (1995–1996). Utrecht: Sardes.

Fukkink, R. (1996). "Peer tutoring" in het leesonderwijs: verslag van het eerste experimentele jaar van Stap Door! Spiegel, 14, 3, 47–71.

Goldstein, H. (1995). Multilevel Statistical Models, 2nd ed. London: Edward Arnold.

Goldstein, H., Rasbash, J., Plewis, I., Draper, D., Browne, W., Yang, M., Woodhouse, G., and Healy, M. (1998). A User's Guide to MlwiN. London: Institute of Education, University of London.

Guilford, J.P. (1954). Psychometric Methods. New York: McGraw.

Holoday, B. and Turner-Henson, A. (1989). Response Effects in Surveys with School-age Children. Nursing Research (methodology corner), 38, 248–250.

Hox, Joop J. (1995). Applied Multilevel Modeling. Amsterdam: TT-Publikaties.

Knäuper, B., Belli, R.F., Hill, D.H., and Herzog, A.R. (1997). Question Difficulty and Respondent Cognitive Ability: The Impact on Data Quality, Journal of Official Statistics, 181–199.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. Applied Cognitive Psychology, 5, 213–236.

Krosnick, J.A. and Alwin, D. (1987). An Evaluation of Cognitive Theory of Response-order Effects in Survey Measurement. Public Opinion Quarterly, 51, 2, 201–219.

Krosnick, J.A. and Fabrigar, L.R. (1997). Designing Rating Scaling for Effective Measurement in Surveys. In Survey Measurement and Process Quality, eds. L. Lyberg, P.P. Biemer, M. Collins, C. Dippo, E. de Leeuw, N. Schwarz, and D. Trewin, New York: John Wiley & Sons, Inc.

Krosnick, J.A., Narayan, S., and Smith, W.R. (1996), Satisficing in Surveys: Initial Evidence. New Directions for Evaluation, 70, 29–43.

Lagerweij, N.W. (1995). Milieugedrag bij kinderen. Ph.D. Dissertation, Universiteit van Amsterdam.

Leigh, J.H. and Martin, C.R. (1987). Do-not-know Item Nonresponse in Telephone Surveys: Effects of Question Form and Respondent Characteristics. Journal of Marketing Research, 24, 418–424.

Miller, G.A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review, 63, 2, 81–97.

Orwin, R.G. (1994). Evaluation Coding Decisions. In The Handbook of Research Synthesis, eds. H. Cooper and L.V. Hedges, New York: Russel Sage Foundation.

Peetsma, T.T.D. (1992). Toekomst als motor? Toekomstperspectieven van leerlingen in het voortgezet onderwijs en hun inzet voor school. Ph.D. Dissertation, Universiteit van Amsterdam.

Rodgers, W.L., Andrews, F.M., and Herzog, A.R. (1989). Quality of Survey Measures: A Structural Modeling Approach. Paper.

Rodgers, W.L., Herzog, A.R. and Andrews, F.M. (1988). Interviewing Older Adults: Validity of Self-Reports of Satisfaction. Psychology and Aging, 3, 1, 264–272.

Schoonen, R., Triescscheijn, B., Van Gelderen, A., and De Klerk, A. (1993). Evaluatie van de leeskisten Het Plein en De Klas van Meester Ed. De inhoud, het gebruik, de leerkrachten en de leesvaardigheid van de leerlingen. Amsterdam: SCO.

Schwarz, N., Knäuper, B., and Park, D. (1998). Cognition, Aging, and Self-Reports. Philadelphia, PA: Psychology Press.

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, 331–350, New York: Wiley.

Stock, W.A. (1994). Systematic Coding for Research Synthesis. In The Handbook of Research Synthesis, eds. H. Cooper and L.V. Hedges, New York: Russel Sage Foundation.

Sudman, S., Bradburn, N.M., and Schwarz, N. (1996). Thinking about Answers. The Application of Cognitive Processes to Survey Methodology. San Fransisco: Josey-Bass.

Tourangeau, R. and Rasinski, K.A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. Psychological Bulletin, 103, 299–314.

Received February 2000
Revised May 2001