

## RESEARCH NOTE

### RESPONSE QUALITY IN SURVEY RESEARCH WITH CHILDREN AND ADOLESCENTS: THE EFFECT OF LABELED RESPONSE OPTIONS AND VAGUE QUANTIFIERS

*Natacha Borgers, Joop Hox, and Dirk Sikkel*

Instead of using their parents or caretakers as informants, children are increasingly the principal informants about their own experiences, perspectives, attitudes, and behaviors (Scott, 1997). As a consequence, it is becoming an important question how well one can ask survey questions of young children and adolescents. Concerning adults, there is a substantial body of empirical evidence that shows effects of both respondent characteristics (especially cognitive abilities) and question characteristics (especially question difficulty) on response quality (e.g. Alwin & Krosnick, 1991; Krosnick & Fabrigar, 1997; Narayan & Krosnick, 1996; Schwarz & Hippler, 1995; Schwarz & Knäuper, 1999; Schwarz, Park, Knäuper, & Sudman, 1998). Because children are still developing their cognitive and social skills, the quality of their responses to survey questions is of special interest. For example, earlier studies on surveying children have shown that vague and ambiguous words should be avoided (Borgers, 2002; Borgers & Hox, 2001; De Leeuw & Otter, 1995), and that completely labeled response options helped children to produce more reliable responses (Borgers, 2002).

#### CHILDREN'S COGNITIVE CAPABILITIES AND SURVEYING

The theoretical background of our study is twofold: Krosnick's (1991) satisficing theory and Piaget's theory of cognitive development. Krosnick's satisficing theory offers an explanation for differences in reliability of responses between respondents, and for the effects of question wording. The satisficing theory elaborates on a standard question-answering process-model (Tourangeau & Rasinski, 1988) that distinguishes four steps: (1) understanding and interpreting the question; (2) retrieving information from memory; (3) making a summarized judgment; (4) reporting this judgment. The satisficing theory identifies two response processes: optimizing and satisficing.

The authors thank CentERdata for the data collection. We also thank Edith de Leeuw and Astrid Stribos-Smits for their readiness to provide material that was of importance for this study, Mick Cooper and Jon Krosnick for their helpful suggestions on earlier versions, and three anonymous reviewers for their supportive comments. Finally we thank the Netherlands Organization for Scientific Research (NWO) for the award of a travel grant to the first author.

The note was first submitted to IJPOR April 9, 2002. The final version was received August 29, 2002.

Optimizing means that the respondent goes through all of the four cognitive steps needed to answer a survey question. Satisficing means that a respondent gives a more or less superficial response that appears reasonable or acceptable, without going through all the steps involved in the question-answering process. Satisficing is related to three dimensions of the question-answering process: the motivation of the respondent, the difficulty of the task, and the cognitive abilities of the respondent. Low motivation, difficult questions, and low cognitive abilities may lead respondents to providing a satisfactory response instead of an optimal one. Implicitly, the satisficing theory assumes an interaction effect between respondent characteristics and question characteristics, because the less cognitively sophisticated respondents are more sensitive to difficult or cognitively demanding questions.

Although Piaget's theory of cognitive development has received much criticism, his work still is the most comprehensive theory on the nature and development of cognition. The critique focuses mostly on the precise timing of the successive stages, also the transitions from one stage to the next are not as clear-cut as assumed. In addition, Piaget tends to underestimate the abilities of young children and overestimates those of adolescents and adults (Gray, 2002). Despite its shortcomings, Piaget's theory of successive stages in children's cognitive development provides the impetus for much psychological research and presents a useful framework for practical applications (Flavell, Miller, & Miller, 1993). In our research, the theory enables us to combine developmental abilities and the cognitive demands of survey research. Combined with the question-answering process and the satisficing theory, it helps explain why young children may have more difficulties with cognitively demanding survey questions than older children.

Piaget's cognitive developmental theory distinguishes five stages (Flavell, 1985; Gray, 2002; Inhelder & Piaget, 1958; Piaget, 1929). Until the age of 7 children develop the basic skills necessary for successful verbal exchange. The age group as a whole is limited in their language development, which implies strong limitations in comprehension and in verbal memory. According to Piaget (1929) the age of 7 is a major cognitive turning point. In the fourth stage (concrete operations), children aged 7 until 11 develop language skills and acquire reading skills. However, children of this age still have problems with logic, for instance negations, and tend to be very literal in the interpretation of words (Flavell et al., 1993). In agreement with Piaget's view that young children have problems with logical negations and abstract thought, Holaday and Turner-Henson (1989) found that children have difficulties with ambiguous and vague words because they tend to interpret the words literally. For instance, offering vague quantifiers in questions about the frequency of behavior produces difficulties for children because they need clear definitions (Borgers & Hox, 2000; De Leeuw & Otter, 1995; Holaday & Turner-Henson, 1989; Robinson, 1986). Another example is partially labeled options, which lack clear definitions of the offered response options. Since children of these ages are limited in their logical and abstract thinking it will be very difficult for them to cope with partially labeled response options, as has been found before (Borgers, 2002), because they have to interpret and translate the unlabeled options themselves.

In the fifth stage (formal thought), with children aged 11 until 15, cognitive functioning is well developed, including formal thinking, negations, and logic. Piaget's description of adolescent thinking abilities has proven to be quite accurate (Keating, 1980; Niemark,

1982). Children in this age group are able to manipulate ideas about hypothetical situations (Conger & Galambos, 1996). However, they are very context sensitive and they may have their own norms. Adolescents aged 16 and up are generally treated as adults in survey research.

Since answering survey questions seems to be a more difficult task for children than for adults, one may expect that they have to resort to a satisficing strategy more often. The precise effect of satisficing is unclear. To the extent that satisficing leads to more error in the response process, it should result in less reliable responses than using an optimizing strategy. To the extent that satisficing implies a simple but systematic heuristic in the response process (e.g. always give the most socially desirable answer), it may result in an apparent increase in the reliability of the responses, of course without increasing the validity. In this note, we present the results of an experiment in which questions were administered in different formats: using clear versus vague quantifiers and using completely or partially labeled response options. One indicator of response quality we use is stability over time. Offering vague quantifiers and partially labeled response options should produce less stable responses because these questions require logical and abstract thinking, which should lead to more variable responses over time (De Leeuw & Otter, 1995; Vaillancourt, 1973). In addition these formats should be more difficult to cope with for the younger children (before the formal thought stage).

In addition to the effect of these question characteristics on the stability of responses, we investigate whether different question wording affects the multivariate relationships between variables. It is possible that changes in the response scale only shift the scale, or induce additional error, without changing the pattern of relationships with other variables. This has been called the 'form-resistant correlation' hypothesis (De Leeuw, 1992; Krosnick & Alwin, 1987). However, the alternative hypothesis is that changes in the question-response process do in fact lead to differences in the measurement process that affect the relationships between the variables we attempt to measure.

With respect to the cognitive developmental stages of children and the influence of the cognitive abilities on the question answering process, we formulate four hypotheses for this study:

- (1) Partially labeled response options and vague quantifiers produce less stable responses compared to the completely labeled response options and the use of clearly quantified words in the response options.
- (2) The younger the children are, the less stable their responses will be.
- (3) The younger children are, the larger the effect of partially labeled response options and offering vague quantifiers will be.
- (4) The relationship between variables differs for the questions with vague and labeled response options, vague and partially labeled response options, and clear and completely labeled response options.

## METHOD

### DATA COLLECTION

CentERdata, a University-based research center that operates a telepanel, has collected the data. The CentERdata telepanel consists of a random sample of approximately 2000

Dutch households who have an Internet connection at home. The panel members receive questionnaires via Internet on a weekly basis. New and replacement households are selected by telephone interviews (CATI). The sample for these interviews is based on random selection of telephone number of Dutch private households (Felix & Sikkel, 1999). Panel members who do not have a computer or an Internet connection make use of a set-top box from CentERdata, which enables them to use e-mail and the Internet and to fill out electronic questionnaires. The panel is representative of the Dutch population with respect to age, gender, religion, education, and region (CentERdata, 2002).<sup>1</sup>

For this experiment all children in the households who were between 8 and 16 years old (means = 11.6; *s.d.* = 2.19), were asked to answer the questionnaire. The questionnaire was administered twice (repeated measures). The first administration took place between the end of June and the end of September. In the first administration 222 children participated, of whom 117 are boys and 105 are girls. In the second administration only 91 children responded. The period between the first and second measurement varies between three and eight weeks (median = 4). The enormous loss of respondents in the second measurement occasion can partially be explained by the period of data collection, the summer holiday. Although the response dropped by more than half of the respondents, there are no significant differences in responses between both measurement occasions on all questions used in this experiment.<sup>2</sup> In addition the three (randomly assigned) experimental groups did not differ significantly ( $p < .05$ ) in mean age of the children, or in their attrition rates.

## DESIGN

The experiment was designed with three different response formats: completely labeled—vague response options; partially labeled—vague response options; completely labeled—clear response options. The Appendix presents the response options as offered in the questionnaire.

Three instruments were used in this study. The first two instruments ask how often children carry out several activities (15 questions) and how often they read several kinds of books (14 questions). The third instrument contains a scale of 5 questions about parental educational stimulation. All questions used a 5-point response scale. The question order (per instrument) was randomly assigned to the children.

<sup>1</sup> The data are collected in a standing household panel. The parents in the household receive the questionnaires and the information that the children in the household between the ages of 8 and 18 years should reply to this questionnaire. At the moment of data collection the panel was in transition. CentERdata changed their system, which caused many problems. One of the problems was that many panel members (households) quit the panel and were replaced by others that were included in the panel. As a result, for this specific time period there is no information on how many households were included in the panel. Especially the number of children in the households for that period is unknown. As a result the precise response ratio cannot be calculated.

<sup>2</sup> Differences were tested with *t*-test for all variables in our experiment, using a Bonferroni correction for  $\alpha$  ( $\alpha = \alpha/k = 0.001$ ). There were no significant differences between both measurement occasions.

## ANALYSIS AND RESULTS

## STABILITY OF RESPONSES OVER TIME

To test the first three hypotheses a measure for stability of responses over time is used as dependent variable. In total 34 questions were answered by 91 children on two measurement occasions. Per child and per question the absolute difference between both measurements is taken as an indicator of the stability of the response over time.

The data can be viewed as a hierarchical data set with questions nested within children. To analyze such data a multilevel regression model is appropriate. Because of the strongly skewed distribution of the response variable and the relatively small sample size ( $N=91$ ), a nonparametric bootstrap method was used to validate the statistical results ( $p$ -values) of the final model.<sup>3</sup>

Table 1 shows the results of the multilevel regression analysis of the effects of the child characteristics and question characteristics on the stability of responses. The first column presents the results of the so-called null model and the last two columns present the bootstrapped estimates and the 95 percent bootstrap-confidence interval. The null-model decomposes the variance in the dependent variable over the two levels. This shows that only 5 percent of the total variance (.03/.56) in the stability over time is ascribed to the child's level while 95 percent (.53/.56) is ascribed to the question level. This means that the variance of the absolute difference over time depends mainly on the differences between questions and only for a very small part on the differences between children.

Table 1 shows that the child characteristics, age, gender, and the number of weeks between the first and second measurement, did not show significant effects in the final model. The absence of an effect for the number of weeks between the two measurement occasions confirms that there are no direct memory effects involved, which is satisfactory. The manipulated question characteristics do not show a significant effect either. The only significant effect is found for the interaction between age of the children and labeled response options. This interaction effect replaces the significant effects of age ( $r_{xy} = -.02$ ;  $s.e. = .01$ ) and labeled response options ( $r_{xy} = -.13$ ;  $s.e. = .05$ ) themselves. Thus, the effect of offering completely labeled responses exists only in the older age groups. However, regression coefficients that are part of an interaction cannot be interpreted separately (Aiken & West, 1991). Figure 1 presents the interaction in graphical form. It clearly shows that completely labeled response options have an effect only with the older children. Contrary to our expectations, the youngest children do not show differences between both offered options. The older children definitely do better with completely labeled response options.

## EQUIVALENCE OF MEASUREMENT

To test the hypothesis of factorial equivalence, a structural equation model has been analyzed using AMOS (Arbuckle & Wothke, 1999). For this, all available data from

<sup>3</sup> In the nonparametric bootstrap the residuals from the multilevel regression model are resampled in each bootstrap iteration (Hox, 2002). This produces parameter estimates and standard errors that are robust against violations of distributional assumptions and small sample sizes.

TABLE 1 Effect of child and question characteristics on the stability over time (Multilevel regression analysis)

	Null-model	Results complete model ( $N=91$ )	
	parameter (s.e.)	Bootstrapped estimates	95 percent- confidence interval
Fixed part			
Intercept	<b>.53 (.02)</b>	<b>.41</b>	<b>.00 / .80</b>
<i>Child characteristics</i>			
Age		.01	-.03 / .04
Gender		.02	-.08 / .11
Weeks		.02	-.01 / .04
<i>Question characteristics</i>			
Vague response options		.04	-.06 / .14
Completely labeled response scale		.37	-.12 / .89
<i>Cross level interaction</i>			
Age $\times$ Labels		<b>-.04</b>	<b>-.08 / -.00</b>
Random part			
$\sigma^2$ child level	<b>.03 (.01)</b>	<b>.02</b>	<b>.01 / .03</b>
$\sigma^2$ item level	<b>.53 (.01)</b>	<b>.53</b>	<b>.50 / .56</b>

Note: significant results ( $\alpha=.05$ ) are printed bold

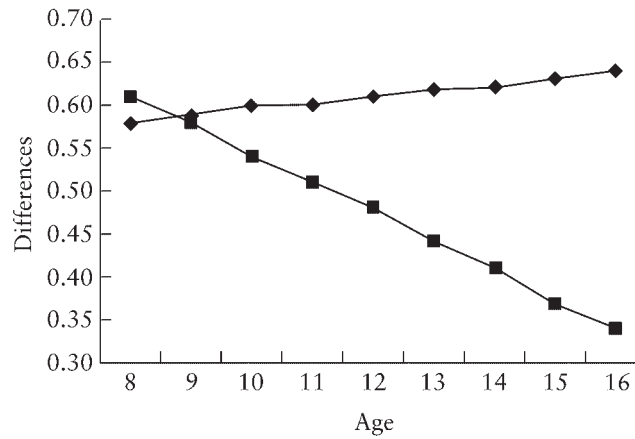


FIGURE 1 Interaction effect between age of children and labeled response options on the difference between both measures

the first measurement occasion were used ( $N=222$ ). First, an acceptable and well-fitting model was estimated on the whole data set, disregarding the differences in the offered response options. The model includes three factors relating to the types of books the

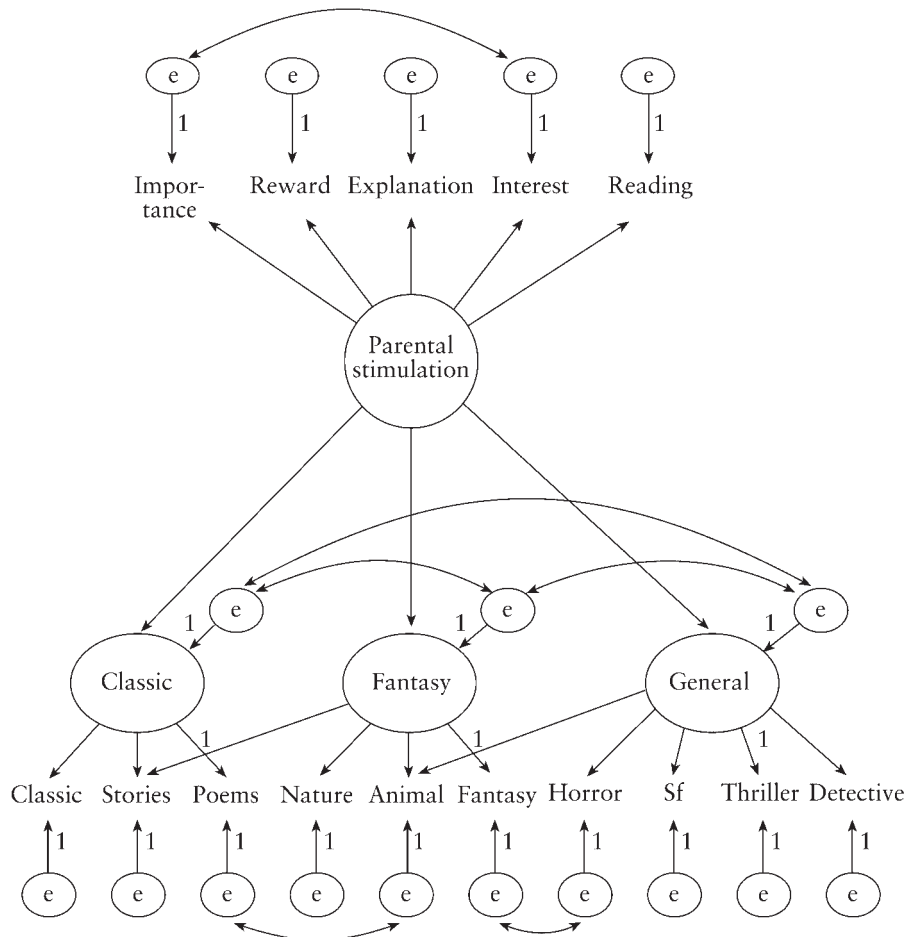


FIGURE 2 Base model for testing equivalence of measurement between three groups

children are reading, which are influenced by the amount of parents' stimulation concerning school and reading activities. Figure 2 presents the model in graphical form.

This was followed by a multiple group analysis to test for factorial equivalence between the three experimental groups: (1) completely labeled vague response options; (2) partially labeled vague response options; (3) completely labeled clear response options. We started with the strictest model in which each parameter, as specified in the model presented in Figure 2, is assumed to be equivalent across all three groups. In the second model equality restrictions between the three groups are still imposed on the factor loadings and correlations between the three error components of the three latent variables, but differences were allowed in the relations between the latent variables. In the last model all parameters are allowed to be different between the three groups. The only restriction left is the structure of the model itself.

The fit of the model as shown in Figure 2, estimated on the whole data set disregarding

differences in the response options is reasonable ( $\chi^2=90.9$ ;  $df=79$ ;  $p=.169$ ). The strictest model, in which each parameter is assumed equivalent across the three groups is rejected by the chi-square model test ( $\chi^2=445.7$ ;  $df=319$ ;  $p=.00$ ) and by unacceptable goodness-of-fit indices (Tucker Lewis, GFI, and AGFI all  $<.80$ ). The model in which the equality restrictions between the three groups are still imposed on the factor loadings and correlations between the three error components of the three latent variables is also rejected by the chi-square model test ( $\chi^2=356.5$ ;  $df=275$ ;  $p=.00$ ), and the goodness-of-fit indices (Tucker Lewis, GFI, and AGFI all  $<.85$ ). A last model in which all parameters were allowed to be different across the three groups was not identified.<sup>4</sup>

Although the model with partial restrictions is by itself not acceptable, it fits significantly better than the model with all restrictions ( $p < 0.00$ ). The last model in which all parameters are allowed to be free was not identified. However, we can identify the multivariate differences between the three experimental groups by inspecting the modification indices, which are suggestions to improve the fit of the model. This helps us to gauge the magnitude of the multivariate differences between the different question formats. The questions with completely labeled clear response options do not show large modification indices, which indicates a good model fit in this group. For the other two formats, the large modification indices indicate both correlated measurement errors and some direct paths between observed variables.<sup>5</sup> Both modifications strongly suggest that changing the response options by making them less informative has a negative impact on the measurement part of the model. This supports our last hypothesis that multivariate relationships between variables can indeed be affected by changes in the response format of the questions, because these changes harm the measurement process.

## CONCLUSION

This study focuses attention on four hypotheses: first the effect of child characteristics and offering vague quantifiers and labeled response options on the stability over time was tested; as well as the equivalence of measurement across the three question formats.

With respect to the first three hypotheses concerning the effects on stability over time, we found some unequivocal results that are contrary to our expectations. First, in contrast with earlier studies and our expectation, this study does not find an effect of offering vague quantifiers in the response options on the stability of responses over time. One possible explanation could be the use of CAPI with a standing telepanel. The use of computer assisted questionnaires in general produces a better response quality in survey research with children (De Leeuw, Hox, Kef, & Van Hattum, 1997; Van Hattum & De Leeuw, 1999).

In general older children produce smaller differences over time, as do completely labeled response options. However, these main effects disappear after including the

<sup>4</sup> The model meets formal criteria for identifiability; we attribute the convergence problem to the small sample sizes in the separate groups analysis.

<sup>5</sup> For questions with partially labeled vague response options three correlated errors should be included as well as a path from interest to poems and a path from poems to interest. For completely labeled vague response options four correlated should be included as well as a path from reward to importance, one from importance to reward and from stories to horror.



interaction effect between age and labeled response options. The positive effect of offering completely labeled response options only remains for children above the age of 10 or 11. Contrary to our expectations it is not true that younger children have more difficulties with cognitively demanding questions. All age groups are equally affected by the more difficult question format, partially labeled response options. Conversely, young children do not benefit from the extra information that is offered in the completely labeled response options. The improvement of stability over time only appears with the older children. The younger children produce responses with a certain amount of error, which is stable across the different conditions. Apparently, for them *all* questions are difficult questions. Older children can take advantage of the completely labeled response options, because their abilities make it possible for them to understand and interpret all given information in the question. As mentioned in the introduction, around the age of 11 children develop formal thinking, which is necessary for an optimal question answering process. Recognizing and benefiting from improvements in the response format requires cognitive processes, which are not yet developed within children before the age of 10 or 11.

With respect to the second research question we conclude that offering different types of response options can indeed lead to substantively different structural models. The three experimental groups do not show measurement equivalence. Thus, the observed differences do not just reflect a shift of the position on a specific variable, and the 'form resistant correlation' hypothesis is rejected. It should be noted that De Leeuw (1992) who investigated the 'form resistant correlation' hypothesis in the context of mode effects also rejects it. Generally speaking, it appears to be that offering the clearest type of response options produces the best data quality in questionnaire research with children.

Established research with adults shows many effects of all kinds of question characteristics on response quality. Given our results, we question if all these effects are transferable to children. We assume that the effects of question characteristics found with adults will mostly be absent with children before the age of 10 or 11, because young children cannot fully apply an optimizing strategy and as a result will not benefit from improvements in question formats.

#### APPENDIX: RESPONSE OPTIONS OFFERED IN THE QUESTIONNAIRE

- Completely labeled vague response options: 1 = never  
 2 = rarely  
 3 = sometimes  
 4 = often  
 5 = very often
- Partially labeled vague response options: 1 = never  
 2  
 3  
 4  
 5 = very often



- Holaday, B., & Turner-Henson, A. (1989). Response effects in surveys with school-age children. *Nursing Research*, 38 (4), 248–250.
- Hox, J. J. (2002). *Multilevel analysis. Techniques and applications*. Mahwah, NJ: Erlbaum.
- Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.
- Kaplan, D. (2000) *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Keating, D. P. (1980). Thinking processes in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology*. New York: Wiley.
- 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., & Alwin, D. F. (1987). An evaluation of cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51, 201–219.
- Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scaling for effective measurement in surveys. In L. E. Lyberg, P. P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 141–164). New York: Wiley & Sons.
- Narayan, S., & Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, 60, 58–88.
- Niemark, E. D. (1982). Adolescent thought: Transition to formal operations. In B. B. Wolman & G. Strickler (Eds.), *Handbook of developmental psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Piaget, J. (1929). *Introduction to the child's conception of the world*. New York: Harcourt.
- Robinson, W. P. (1986). Children's understanding of the distinction between messages and meanings: emergence and implications. In M. L. Richards (Ed.), *Children of social worlds*. Cambridge: Polity Press.
- Schwarz, N., & Hippler, H. J. (1995). Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opinion Quarterly*, 59, 93–97.
- Schwarz, N., & Knäuper, B. (1999). Cognitive aging and accuracy of self-report. In D. Park & N. Schwarz (Eds.), *Cognitive aging: a primer* (pp. 233–252). Philadelphia, PA: Psychology Press.
- Schwarz, N., Park, D., Knäuper, B., & Sudman, S. (1998). *Cognition, aging, and self-reports*. PA: Psychology Press.
- Scott, J. (1997). Children as respondents: methods for improving data quality. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurements and process quality* (pp. 331–352). New York: Wiley.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299–314.
- Vaillancourt, P. M. (1973). Stability of children's survey responses. *Public Opinion Quarterly*, 373–387.
- Van Hattum, M. J. C., & De Leeuw, E. D. (1999). A disk by mail survey of pupils in primary schools: Data quality and logistics. *Journal of Official Statistics*, 15, 413–429.

### BIOGRAPHICAL NOTES

Natacha Borgers is a teacher at the department of social science methodology and statistics at Utrecht University, the Netherlands. She is currently finishing her dissertation, which is about response quality in questionnaire research with children.

Joop Hox is Professor of Social Science Methodology at the department of Methodology and Statistics of the Faculty of Social Sciences at Utrecht University, and also teaches at the International University College, Utrecht. His research interests focus on two lines of work: data quality in social surveys and multilevel modeling.

Dirk Sikkel worked as senior statistician at Statistics Netherlands. After that he was research manager in commercial marketing research. In 1993 he became professor for Modeling in the Social Sciences at the University of Amsterdam. In 1996 he moved to Tilburg University to become professor for Data Collection in the Economic Sciences.

Address correspondence to Natacha Borgers, University of Utrecht, Methodology & Statistics Department, P.O. Box 80.140, NL-3580 TC Utrecht, The Netherlands, Email: [n.borgers@fss.uu.nl](mailto:n.borgers@fss.uu.nl)