Improving Data Quality when Surveying Children and Adolescents:

Cognitive and Social Development and its Role in Questionnaire Construction and Pretesting

Prof. Dr. Edith D. de Leeuw

Department of Methodology and Statistics, Utrecht University (email: *e.d.deleeuw@uu.nl)*

Acknowledgements

Contact Address:

Professor Dr. Edith. D. de Leeuw

Plantage Doklaan 40

Nl-1018 CN Amsterdam, The Netherlands

e-mail: *edithl@xs4all.nl*

Content

Preface

In this report, we integrate the current theoretical and empirical knowledge regarding questionnaire research with children and adolescents, focusing on how children in successive stages of cognitive and social development differ from adult respondents.

In the first part, we formulate rules for the construction of questionnaires for different age groups and give examples of reformulation of questions tailored to different age groups.

In the second part, we focus on methods for pretesting questionnaires. Different methods can and should be used for different age groups.

In the last part, we present a checklist for questionnaires directed at children and adolescents. This checklist can be used both as a guideline when constructing a questionnaire, and as a coding scheme for the evaluation of questionnaires (e.g., expert-evaluation).

This report is based on earlier publications on surveying children and adolescents.

Key references are:

1. De Leeuw, E.D. (2005). Surveying Children. In S.J. Best and B. Radcliff (Eds). Polling America: An Encyclopedia of Public Opinion. Westport. CT. Greenwood Press, pp. 831-835.
2. Edith D. de Leeuw (2003). Questioning Children in Survey, Report to the Statistical Information and Monitoring Programme on Child Labour of the International Labour Office's International Programme on the Elimination of Child Labour Meeting April 14-15 2003 in Geneva.
3. Natacha Borgers (2003) Questioning Children's Responses. Ph.D Thesis, University of Utrecht
4. De Leeuw, E., Borgers, N., & Smits, A. (2004). Pretesting questionnaires for children and adolescents. In: S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, & E. Singer. Methods for testing and evaluating survey questionnaires. New York, Wiley, 2004, chap 20, pp 409-429.

Edith D. de Leeuw. Amsterdam May 2011

Executive Summary

Definition of survey according to the Section on Survey Research Methods of the American Statistical Association (www.amstat.org/sections/srms): A survey is a method of gathering information from a sample of individuals. Different data collection methods may be used (e.g. self-administered questionnaires, internet or postal mail methods, or telephone and personal interviews), but always a structured well-defined questionnaire is used.

From which age can a child be surveyed? Children form a special population, and children from 7 to 18 have to be surveyed in special surveys. Below the age of 7 children do not have sufficient cognitive skills to be effectively and systematically questioned. The age of 7 is a major developmental point in the cognitive and social maturation of children, and with care children can be interviewed with structured questionnaires or complete self-reports from 7 years onwards, depending on their development. At the age of 18, adolescents are generally treated as adults in surveys, as is reflected in definitions of adult populations for many surveys (De Leeuw, 2003; De Leeuw, Borgers & Smits, 2004). However, not all children have the same chances and opportunities. The age of 7 is mainly based on studies in Western-Europe and the United States, and even in these privileged circumstances not all children develop equally fast. Children in less privileged circumstances (lack of schooling, malnutrition, etc) may develop slower.

When use proxy reports instead of self-reports. Children older than 7-10 may be surveyed directly, and the older the child, the more reliable the answer will be. A good rule is to try and collect information directly from children on topics for which they are the best informant, such as their feelings, and other subjective phenomena. Children are also the best respondents on factual or general questions that are outside the scope of parents' or guardian's knowledge. For example, working conditions. An example from the USA is food intake and eating habits (as children often do not eat the food taken with them to school). Till adolescence, a well-informed adult proxy will likely provide better data in all other cases. For example, on facts about schooling, family, and health related issues, such as visits to the doctor and vaccinations, an informed parent will have more accurate knowledge. Below the age of 7, direct questionnaire research of children is not feasible at all, between 7-10 careful pretesting

should always be done, to decide if direct questioning is feasible. In many cases, an informed parent or daily caretaker can serve as a proxy respondent and provide information on daily activities, health issues, or other topics of interest.

Different questions for different age groups: Young respondents between the age of 7 and 18 are a far from homogenous group. As children grow from infancy to adulthood, their thinking becomes more logical, and their reasoning skills develop more and more. At the same time memory and language develop and social skills are acquired. These are important prerequisites for the understanding of survey tasks and questions. Children's questionnaires should be tailored according to the cognitive and social development of the intended age group. This means that questions for young children should be simple both in question structure and question wording. Globally three relevant age-groups can be discerned: middle childhood, early adolescence, late adolescence.

Middle childhood (7-12): *Language skills* are sufficiently developed to use individual semi-structured interviews. From around 9, structured interviews are feasible. Children of this age can answer well-designed questions with some consistency. But, as reading and language skills are still developing in middle childhood, the understanding of words has to be checked very carefully for this group. Questions using logical operators such as 'or' (e.g., does your father or mother…) or negations are *not* yet understood correctly. Extra attention should be paid to complexity of wording. Questions should consist of several short sentences, in tead of one long sentence. As children in this age group can be very literal, depersonalized or indirect questions should be checked very carefully. When preparing the questionnaire, one should take care that both questions and instructions are simple, and that the *question wording* is clear and unambiguous.

Memory and cognitive processing time is a second important issue. In middle childhood (7 to 12) both memory capacity and memory speed is still developing. Therefore, complexity of the question and number of response categories should be carefully examined. If possible, one should use visual stimuli and response cards, to make the task more concrete and interesting. Response cards are very helpful, as young children tend to forget even a limited set of response options and have difficulties with more than 2-3 verbal response categories. Retrospective questions may pose extra problems, and young children are prone to construct scripts of familiar routines if they do not clearly recollect events.

Emotional and social development of the intended age group should be taken into account. In younger children suggestibility is an important item. In early middle childhood (7 to 10) children have a tendency to please and are afraid of doing something wrong. This may result in more superficial answers and in an inclination towards social desirability.

Early adolescence (12 to 16)*: Cognitive functioning* is now well developed, and logical operators (e.g., and, or) and negations (e.g., not) are understood. It is possible to use standardized questionnaires similar to questionnaires for adults, but one should guard extra against ambiguity of *question wording*. Keep questions as simple and as concrete as possible. Memory capacity is now full-grown, but memory speed is not. Even in this older age group ample time for answering questions should be allowed.

From the age of 12, peers become increasingly important, making adolescents increasingly sensitive to *peer pressure and group norms*. As a result, sensitivity of topic and *privacy of interview situation* become important. Both the nearness of schoolmates or siblings and parents can influence the answers dramatically. Ensuring privacy, and stating confidentiality is important. More confidential methods, such as self-completed questionnaires may also help to obtain better answers. But above all, one should keep this group motivated and guard against boredom.

Late adolescence: from 16 years onwards, adolescents may (with care) be regarded as adults with respect to cognitive development and information processing. However, the adolescent brain is not yet fully developed and especially functions with regard to organizing and social skills are not yet completely developed. Furtehrmore, resistance to peer pressure is still very low and older adolescents have their own group norms and social norms. The social context of the survey (e.g., presence of siblings or friends) remains extremely important, especially in interaction with special topics.

Informed consent: ESOMAR (www.esomar.org), the world association for research professionals in opinion polling and market research, gives explicit guidelines. ESOMAR states that first of all a researcher should conform to any relevant definitions in any national code of conduct and/or in national legislation, and second that in the case of children under 14 explicit permission should be asked of a parent, guardian, or other person the parent has conferred responsibility to. Of course, national legislations may differ regarding the age at

which children can legally give their consent. However, permission of a parent or guardian is not enough. Professional research organizations like the Society for Research in Child Development require that researchers inform the child about the study and obtain permission of the child in addition to the consent of the legal guardian. This implies that the information presented to the child should be given in clear language and at a level that the child can understand.

In sum: Special care should be given to the construction of questionnaires for children and adolescents. One and the same questionnaire for the total age group 7-18 is not recommendable: questionnaires should be tailored to the cognitive and social maturity of the child. In addition, pretesting of the questionnaire is necessary to examine the adequacy of *question wording* and response options for different age groups. On may also pretest other aspects of the questionnaire and of the survey design. For example, do children in the relevant age group clearly understand general instructions and introductions to questions? Do they understand the task that is being asked of them? Do they understand the request for *informed consent*?

References:

Edith D. de Leeuw (2003). Questioning Children in Survey, Report to the Statistical Information and Monitoring Programme on Child Labour of the International Labour Office's International Programme on the Elimination of Child Labour
Meeting April 14-15 2003 in Geneva.

Edith de Leeuw, Natacha Borgers, and Astrid Smits. "Pretesting questionnaires for children and adolescents." In: S. Presser, J. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds). Methods for testing and evaluating survey questionnaires (2004). New York: Wiley.

Introduction

When constructing and evaluating questionnaires for children, a researcher should start with following the basic rules for general questionnaire construction as outlined in handbooks such as Dillman (1978, 2000), and Fowler (1995). These include the good advice to use simple words, avoid ambiguity, ask one question at a time, etc. But, one has to do more. Methodological studies on adult populations have shown that adults may experience problems with certain questions, and that question characteristics may affect the data quality in surveys. With children as respondents, the same problems are magnified, as an error (e.g. ambiguity, negations) in the questionnaire is more difficult to compensate, or has a larger impact. In addition to a magnification of the problems experienced by adults, children experience specific problems when responding. These specific age-dependent problems can be better understood within the conceptual framework of the question-answer process, that is the cognitive process that takes place between the posing of the question and the delivering of an answer (for a discussion of the question-answer process see Tourangeau & Rasinski 1988).

In children the cognitive, communicative and social skills are still developing. These changes in cognitive and communicative functioning have profound implications for the question-answer process, namely (1) question comprehension, (2) recall of relevant information from memory, (3) judgment formation, and (4) reporting. We will describe each stage and its consequences for questionnaire development, resulting in practical rules and recommendations. A comprehensive checklist is presented in Appendix A .

In the second part methods for pretesting questionnaires for children and adolescents are described.

Part 1: Construction of Questionnaires

Comprehension and Interpretation of Questions

Comprehension and Structure

Before a child can answer a question, she or he has to understand its meaning. First the semantic, literal meaning and secondly the intended meaning of the researcher (Sudman, Bradburn, & Schwarz, 1996). Thus, the child has to interpret and understand both the question *and* the response task; the latter has to do with understanding what the question asks to do in order to provide an adequate answer.

The first point one should check in a draft questionnaire is whether the structure of the questionnaire facilitates the comprehension and interpretation of the questionnaire. Ordering the questions in short blocks, each block containing only one question-format and each block only addressing one topic, will facilitate comprehension.

To make sure that children understand the task, a simple clear and brief introductory text is necessary to introduce each block of questions stating what the next questions will be about (topic) and what is expected of the child (task). This introduction should not be too long. But, if length increases the clarity of the meaning, a longer introduction is preferred. But keep the sentences short! A longer introduction should not be a complex introduction; many simple short sentences can form a long and clear introduction.

Like the introduction, the questions itself should not be too long. Again if the length of the question increases the clarity of the question meaning, a longer question is preferred. But, careful attention has to be given to the structure of longer questions. The pitfall of using long questions is the inclusion of complex constructions, which distracts children from the intended meaning because it requires more technical reading skills than questions without these constructions. A long question that is built up from short clear sentences will not challenge the reading skills that much and may be used in a positive way.

Language and Readability

The second point to pay special attention to when constructing child questionnaires is the language level. The questionnaire should aim at the language level of the intended population. The following should be kept in mind. A six year old will know about 2600 words actively and 8000 works passively. In the school period this number climbs up to 5000 words actively

at the age of 9. In comparison an average adult actively knows 10 000 words. The development of complex sentences takes place during early middle childhood (7-9 years of age). The conjugation of strong and irregular verbs is completed around 9. In middle childhood (7-12 years), children are immediate readers with a strong grasp of phonetics. They continue to develop their whole word recognition skills through exposure to print, and words that are read many times become recognized by their shape. At the beginning of adolescence (around 12) a child is skilled in both phonetic and orthographic (whole word) skills. Also, beyond 12, children learn more advanced skills such as story grammars, writing and audience, style, and other meta-linguistic skills that go beyond mere word recognition.

A useful tool in evaluating questionnaires is a readability index expressed by age. Readability formulae can certainly not be the sole judge of the suitability of a question text, and many other factors need to be considered (e.g., sentence structure, size of type, lay-out). But, they do distinguish clearly between crisp and extended styles of writing and perform better than subjective assessments by expert judges, such as teachers or linguists who usually underestimate the difficulty of a text by several years (Johnson, 2002). Furthermore, readability indices are easy and efficient and give a quick indication of the difficulty of a text.

For general screening purposes, we advice the Flesch-Kincaid readability formula. This is a US government Department of Defense standard tests and results in US school grades (Johnson, 2002). One needs L, the average sentence length (number of words / number of sentences), and N, the average number of syllables per word (number of syllables/ number of words). The US Grade level is then (L x 0.39) + (N x 11.8) - 15.59. To get the reading age comparable to the grade level add 5. So, Reading Age = (L x 0.39) + (N x 11.8) – 10.59 years. This predicted reading level is the 'break-off' point for a reader of that age, based on a 50% correct answer score. Note, that this is still far from full comprehension! Therefore, for practical use the advice is that questions ought to be about two grades less than the official calculated population grade (or two years less than the official reading age).

Readability is more than vocabulary and decoding, it also encompasses comprehension. This is clearly demonstrated by Borgers, de Leeuw & Hox (1999, 2000), who re-analysed a small, but unique, data set that was collected for the evaluation of a Dutch reading stimulation program. Data on reading attitudes were collected for 443 children from level 4 of Dutch primary schools (age 7-8), using an audio-presented self-administered test. The test consisted of 25 attitude questions, each with two response categories (yes/no). Data collection took place in the classroom, where the questions were read aloud by an instructor, and the children recorded their (yes/no) answers on a self-administered questionnaire. For

each child the scores on a battery of official Dutch educational tests were available, including a vocabulary tests, a reading decoding test, and two reading comprehension tests. Reading ability in general (vocabulary, reading decoding, and reading comprehension) did influence the item nonresponse on the audio-presented attitude questionnaire. Children with low reading scores produced more missing data. How well the children in general can understand what they are reading (reading comprehension) influenced the consistency of their responses on the audio-presented attitude test. The more technical aspects of reading (vocabulary, reading decoding) did not have any effect on consistency. It should be emphasized that the questions on the attitude test were presented to the children auditory (read out aloud) to prevent potential influences of literacy. Still, the scores on reading tests and especially on reading comprehension did influence the data quality on this *auditively* presented attitude test. We assume, that it is *language ability*, and not reading ability that is at work here. Children with more developed language skills understand the questions better (step 1 in the question-answer process), and therefore produce better quality data.

Wording: Intended and Literal Meaning

The wording of a question and the syntax are not only of importance for technical reading capacity (vocabulary, reading decoding), but are of extreme importance for reading comprehension and interpretation as well. So, in the next step of questionnaire development one should check several aspects of question wording. Very important for children is that one should avoid ambiguity at all costs. Empirical studies of questionnaire data showed that children have an extremely low threshold for ambiguity and vagueness in questions and cannot cope with it. Ambiguity in questions leads to lower data quality and there is a clear indication of interaction with age (De Leeuw & Otter, 1995). Double-barreled questions, like ambiguous questions, are difficult for all respondents, even for adults, but for younger children (until late adolescence, e.g. 15-16) they are disastrous (Amato & Ochiltree, 1987).

In agreement with Piaget's early view that young children have problems with logical negations and abstract thought, Holoday and Turner-Henson (1989) found that children have difficulties with 'vague' words because they tend to interpret the words literally. For instance, offering vague quantifiers in questions about the frequency of behaviour produces difficulties for children because they need clear definitions, especially in early middle childhood (7-10). For this age group, simple yes/no questions about doing something are better. This group is also extremely sensitive to the slightest suggestion, and any hint of a suggestively phrased question should be avoided. Finally, negatively formulated questions make the intended

meaning ambiguous for children (as it does for adults), and should always be avoided in children's questionnaires. Especially younger children in middle childhood have many problems with negatively phrased questions. Again there is an interaction with age; older children and adults experience these problems in a lesser degree (Benson & Hocevar, 1985; De Leeuw & Otter, 1995).

Comprehension and interpretation is more than the understanding of the semantic, literal meaning of the question. To understand what is required, a child should also grasp the intended meaning. From the age of seven until 12 year children are still very literal in their interpretation of words (Holoday & Turner-Henson, 1989) As a result, the distance between the intended meaning and the literal meaning of the used words can cause serious problems for children of this age. This is even more pronounced when so called depersonalised or indirect questions are used (Scott, 1997, Scot et al, 1995). A clear illustration is the observation made by Scot et al (1995) during pre-testing that in reaction to questions using the term 'people my age …' some children tried to guess the age of the interviewer before answering!

Information Retrieval from Memory

Memory Capacity

After understanding the question a child has to retrieve the relevant information from memory. During middle childhood (7-12 years of age) memory and especially storage and retrieval of information are still developing (Holoday and Turner-Henson, 1989). Presser, Blair, Mack, Ryan, & Van Dyne (1993) point out that the most consistent finding in the literature is that young children's spontaneous recall is less than adults. Before the age of 11 the memory capacity of children and the constructive processes used are not full grown (Cole & Loftus, 1987; Kail, 1990). Especially in young children (early middle childhood 7-10), it is important that the question is not complex and easily recognizable.

A second point to check for when constructing child questionnaires is the reference period. The reference period should be well defined; Amato & Ochiltree (1987) stress the importance of using 'here and now' questions in middle childhood, as do De Leeuw & Otter (1995). In addition, questions asking for numerical quantities should be carefully screened on intended age groups. In middle childhood meta memory skills are not yet developed and

numeric quantities give severe problems (Gray, 2002); the age of 11 is seen as a turning point in memory capacity when children appear to function as well as adult (Kail, 1990).

Retrospective Questions

Retrospective questions should be checked carefully. A general advice is to avoid retrospective questions as far as possible, when asking for non-standard events and details. If the question is immediately recognizable for children and asks about salient and meaningful experiences (e.g. class outing, visit to paediatrician), even children in early middle childhood (7-10 years) can answer correctly as their memory for salient issues is remarkable (Brainerd & Ornstein, 1991). However, several studies have showed that unreliable responses appear if these children are not involved or interested in the subject (Holoday & Turner-Henson, 1989; Vaillancourt, 1973). Especially the younger children are prone to construct scripts or event representations of familiar routines if they do not clearly recollect *unstandard* events (Brainerd & Ornstein, 1991), or when *more complex* questions are asked (De Leeuw & Otter, 1995). Furthermore, there are developmental differences in reality monitoring, and under certain circumstances young children (early middle childhood) have more difficulties to distinguish between imagined events and those actually perceived (Johnson & Foley, 1984). This is corroborated by Saywitz (1987), who found that 8 and 9 year olds tended to have less complete recall and more embellishments than 11to 12 year olds.

Diaries

When extreme care is taken 'diary'-type of research can be done. Diary methods can especially used to overcome difficulties with questions that require complex memory processes, and cannot be answered from memory directly. For example, when the answer required is composed of several units of information, each has to be retrieved from memory and evaluated. Questions that measure physical time use for behavioural related concepts often request these memory processes. For that reason diary methods are often used to measure these types of concepts (Kalfs, 1993). The diary method minimally appeals to cognitive processes and memory, and the 'here and now' type of question, which is especially appropriate for children (Amato & Ochiltree, 1989), is used. The information requested consists of only a single unit and can be easily recalled, making the process that has to carry out in memory simple for the child. Of course to successfully use diaries basic reading and writing skills are required, but even relatively young children in late middle childhood can use diaries successfully. Otter (1993) showed that the use of the diary method with children, aged

9 years, to measure leisure-time reading yield good response quality, produce reliable and valid data responses. Even when the concepts of interest produce questions that burden children's memory, the use of diary method is by far superior compared to self-administered paper and pencil questionnaires (Otter, 1993). Diaries can be regarded as a series of simple questions. However, structured diaries are also used successfully to collect information about peer interactions of children in their final year at primary school (Ralph, Williams & Campisi, 1997).

Nevertheless, diaries have their disadvantages too, they are time consuming for the respondent and children may easily loose interest (Vaillancourt, 1973) An attractive alternative is the use of electronic diaries, which produce good data quality (Kalfs, 1993). Van Hattum and de Leeuw showed that children are very good in computer-assisted self-interviews (CASI); even children from the age of 8 year successfully completed the questionnaire and enjoyed the process (Van Hattum & de Leeuw, 1999). The use of electronic diaries can motivate children because it gears to their experiences and minimize the burden of memory.

Memory Capacity versus Processing Speed

Finally, it is important is to realize that both working memory capacity and the processing speed are developing with age, but not at the same speed (Gray, 2002). After the age of 12 the memory capacity of children and the constructive processes used are full grown (Cole & Loftus, 1987; Kail, 1990), but even a child of 12 still needs approximately 1.5 times as much time as an adult to process information. Therefore, children should be given ample time to complete questions. Before fielding a questionnaire, it should always be checked if children have enough time for each question

Judgment: Information Integration and Evaluation

Social desirability and Pleasing

In the judgment phase a child has to combine all retrieved information into a preliminary answer and evaluate the answer. In questionnaire research with adults, much attention has been paid to social desirability and its influence on the judgment phase. In young children a second process has to be recognized too. Independent of the sensitivity of a question, young

children in early middle childhood tend to please the researcher or teacher. According to Maccoby & Maccoby (1954), children as old as 8 years will assume that the adult knows everything already. In addition, they are afraid to say something wrong or foolish, especially in a situation that resembles school (Delfos, 2000). As a consequence, young children may react to the demand characteristics of the interview situation by responding in social desirable ways (La Greca, 1990), or fall back on other response strategies such as yes-saying just to please and go along (Maccoby & Maccoby, 1954). To reassure children and to avoid anxiety in young children, one should make sure that the questions do not resemble test items or school questions. Furthermore, in introducing the questionnaire it should be emphasized that this is not a school test, that there are no 'wrong' answers and that the researcher does not know the correct answer.

Self-concept and Peer Pressure

During middle childhood the structure of self-concept changes, and in late middle childhood (10-12) children start comparing themselves with others. During early adolescence (12-16) conformity to peer pressure dramatically increases (Gray, 2002). From approximately 10 years on the effect of peers will be more present, children now fulfill to the cultural rules and the values of society, instead of judge behaviour by their own standards and values. Furthermore, in middle childhood children become aware of the possibilities of putting on a facade and intentionally deceiving others (Selman, 1980). This is clearly illustrated by several methodological studies on children as respondents. Borgers and Hox (2001) reanalysed questionnaire data from five studies and found that on sensitive questions the younger children had lesser item-nonresponse than older children, while on non-sensitive issues this was reversed. Van Hattum and De Leeuw (1999) found that a more private setting (CASI) resulted in fewer social desirable answers for children in late middle childhood. Beebe et al. (1998) discovered that among young high school students even the distance between computers in a computer lab, and thus the closeness of peers, influenced the openness of answers. When reviewing questions for children in late middle childhood and early adolescence, extreme care should be paid to sensitivity of questions and peer norms. Also, when interviewing or administering self-administered questionnaires one should try to make the setting as private as possible.

Reporting: Comprehension and Selection of Response Options

Response Options: Number and Label

In the last step, the reporting phase, two different tasks are performed. The child must understand the presented response options, and select the response that fits the preliminary answer best. When deciding on the response options in a draft questionnaire intended for children, one should realize that developmental differences in both communication and reading skills *and* differences in memory influence the final reporting. This has consequences for the amount of verbal response categories that are optimal for different age groups. For adults, five to seven response categories are generally advised (Krosnick & Fabrigar, 1997). For young children (early middle childhood: 7-10), not more than two or three response categories are a workable number. In late middle childhood and early adolescence four to five is advised, while in late adolescents (16 and older) one can safely use the adult five to seven categories (Borgers & Hox, 2001; Hershey & Hill, 1979; Holoday & Turner-Henson, 1989).

A second check point is the labeling of response categories. Clear labels improve the reliability of answers of adult respondents (Krosnick & Fabrigar, 1997), but labeling is even more crucial for young respondents in whom the logical and systematic thought necessary for interpolation of labels is still developing. Any ambiguity in labeling will negatively influence the data quality; this effect is especially strong for the youngest age group that has the least cognitive sophistication (De Leeuw & Otter, 1995).

An exception to the rules for number of response options and labeling can be made when *graphical* response options are used. Scott, Brynin & Smith (1995) successfully interview young children (late middle childhood) using a show card with seven smiley faces as response scale for happiness. In there study, even the youngest boys (10-11) showed a remarkable understanding of the faces, and all children (even the eldest at 16) enjoyed the task and remained motivated. Kirby, Mann, Petit & Woodhead (2002) successfully used a variety of visual five point scales in an evaluation study among children aged 9-12.

Summary Part 1

In surveying children, language ability is an important issue for the comprehension of questions. *Comprehension* of questions and instructions is the first thing that has to be checked. As reading and language skills are still developing in middle childhood (7 to 12), the understanding of words has to be checked very carefully for this group. Extra attention should be paid to complexity of wording, negations, and logical operators (such as 'and, 'or'). As children can be very literal, depersonalized or indirect questions should be checked very carefully.

Memory and processing time is a second important issue. In middle childhood (7 to 12) both memory capacity and memory speed is still developing. Therefore, complexity of the question and number of response categories should be carefully examined with regard to the developmental stage of the child. Retrospective questions may pose extra problems, and young children are prone to construct scripts of familiar routines if they do not clearly recollect events. In early adolescence (12 to 16) memory capacity is full-grown, but memory speed is not. Even in this older age group ample time for answering questions should be allowed.

In younger children *suggestibility* is an important item. In early middle childhood (7 to 10) children have a tendency to please and are afraid of doing something wrong. This may result in more satisficing strategies and an inclination towards social desirability. In late middle childhood (10 to 12) children become less suggestible, but start to compare themselves with others. From the age of 12, peers become increasingly important, making adolescents increasingly sensitive to peer pressure and group norms. Sensitivity of topic and privacy of interview situation become important.

In this first part, we presented rules for questionnaire development, and a checklist for questionnaires for children and adolescents. Guideline is the question-answer process and its four stages: (1) comprehension and interpretation of the question being asked, (2) retrieval of relevant information from memory, (3) integrating this information into a summarized judgment, and (4) reporting this judgment by translating it to offered response options. Table 1 contains a summary of our checklist according to the relevant steps in the question answer process. For the detailed checklist see Appendix A. This checklist is, of course, a helpful tool when developing questionnaires, but we advise a wider use. This check list may also be used as a tool for formal and informal questionnaire evaluation. An example of informal evaluation is the evaluation of a draft questionnaire by colleagues. An example of formal evaluation is the evaluation of a proposed questionnaire by an advisory committee.

Table 1: Condensed Checklist For Youth Questionnaires by Stage in Question-Answer Process.

Comprehension and interpretation of question	
Length of the introductory text	-> in words
	-> in sentences
Question length	-> in words
	-> in sentences
Ambiguity	-> of the question
	-> of the response scale
Depersonalised or indirect question	
Double barreled question	
Complex construction of question	
Negatively formulated question	
Suggestively phrased	
Retrieving relevant information from memory	
Complexity of the question	
Reference period	
Numerical quantity	
Judging the retrieved information	
Sensitivity of question (question threat)	
Balance of the question	
Position in the questionnaire	
Reporting by communication final response	
Number of response options	
Offering midpoints	
Offering labeled scale points	
Offering Don't know filter	
Offering visual response cards	

Part 2: Pretesting of Questionnaires

Why Pretesting

Designing and conducting quality surveys requires a careful decision process (e.g., Czaja and Blair, 1996; Lyberg, Biemer, Collins, de Leeuw, Dippo, Schwarz, and Trewin, 1997). Designing surveys for children and adolescents is no exception, however with young respondents some design issues are of extreme importance and warrant extra attention. Question wording, structure and length of questionnaire are important factors for data quality as is the developmental phase of the young respondent, both cognitive and emotional/social. When developing and evaluating questionnaires for children, a researcher should start by following the basic rules for general questionnaire construction and evaluation as outlined in handbooks such as Converse and Presser (1986), Dillman (1978, 2000), Foddy (1996), and Fowler (1995). These include good advice to use simple words, avoid ambiguity, ask one question at a time, etc. But, one has to do more. Methodological studies on adult populations have shown that adults sometimes experience problems with certain questions, and that question characteristics affect the data quality in surveys (cf. Krosnick and Fabrigar, 1997). Evidence for interaction effects between respondent characteristics and question characteristics has been found by Borgers and Hox (2001), De Leeuw and Otter (1995), Knäuper, Belli, Hill, and Herzog (1997), and Schwarz, Park, Knäuper, and Sudman (1999). All studies showed that the less cognitively sophisticated respondents are more sensitive to more difficult or cognitive demanding questions than the more cognitively sophisticated respondents, resulting in more item nonresponse and less reliable answers for respondents lower in cognitive ability.

With children as respondents, these problems are magnified. In addition, children experience specific problems when responding. Not only their cognitive, but also their communicative and social skills are still developing, and this affects different stages of the question-answer process, and special care should be given to the construction of questionnaires for children and adolescents.

Therefore, pretesting is a necessary step to take before a survey is fielded (for a comprehensive overview see, Campanelli, 2008). A first step is expert evaluation, in which an expert in both questionnaire construction and the substantive topic of research carefully evaluates the draft questionnaire. The checklist in Appendix A is a helpful tool here.

After consulting with experts in the field, the next step is evaluating the procedures and questionnaire using cognitive testing methods using the intended respondents (children or adolescents of certain age groups) as informants. Cognitive pretests will enable the researcher to discover which wordings or questions are problematic for young respondents and why, thereby suggesting improvements in questionnaires for children.

Cognitive pretesting of the questionnaire is certainly necessary to examine the adequacy of question wording and response options for different age groups. While pretesting theories and procedures for adults are well developed (for an overview, see Campanelli, 2008), for children and adolescents this is still a new field and only few publications about procedures and results are available. Levine and Huberman (2002) describe how they effectively used cognitive interviewing (think-aloud with probing) with children aged 9-14 to test questions from the U.S. National Assessment of Educational Progress. Hess, Rothgeb, and Zukerberg (1998) describe similar positive experiences with adolescents aged 12-17, when pretesting the youth part of the U.S. Survey of Program Dynamics.

In the next section well-known cognitive methods for pretesting with adults (e.g., Esposito and Rothgeb, 1997) are reviewed for usability with children. In addition, we discuss how these methods can be optimized for children.

Focus Groups

Different pretest methods have different strengths (Presser and Blair, 1994). The strength of focus groups is the interaction within the group; the participants stimulate each other to discuss topics and explain ideas (Morgan, 1997). As a consequence a wide range of information can be gathered in a short time, however, this information is not always very detailed. Focus groups are useful to generate ideas and topics for questions, to evaluate the data collection procedures planned, and to evaluate the acceptability or sensitivity of certain topics, but for a detailed evaluation of the questions in-depth interviews are more useful (Snijkers, 2002; Campanelli, 1997).

The usefulness of focus groups in the design phase of a survey is well illustrated by Scott, Brynin, and Smith (1995), who conducted a series of six focus groups with children aged 11-15 in the United Kingdom. The decision to add a 'Young People Survey' to the British Household Panel challenged the researchers to design a way in which children can be interviewed in their homes in privacy. Because of potential literacy problems, the researchers

opted for pre-recorded walkman interviews with a paper self-completion response booklet. Goal of the focus groups was to help develop structured questions and to fine-tune the walkman method. The focus groups took place in a neutral setting, the interviewer's home. Groups were separated by gender and by age groups (11-13 and 13-15) and lasted about two hours with a snack break at half time. Each focus group started with a general open discussion on health and health related issues. This served as warming-up, but also provided information on the typical language use and on sensitivity of topics. This was followed by trying out formats for semi-structured questions thought suitable for these age groups (e.g., response card with range of smiley faces). Question-formats were presented and discussed in the group. In the last phase of the focus group, walkmans were handed out together with a short self-completion booklet. According to the researchers, the focus group discussions were very productive for identifying appropriate wordings, question formats and response options for the development of the Young Person's Questionnaire. The walkman test showed that children did not experience any technical problems when using a walkman and provided useful feedback on voice type (Scott, Brynin, and Smith, 1995; see also Scott, 1997). A subsequent test of the redesigned procedure during the pilot phase of the Young Person's Survey was very successful (Scott, 1997).

Using focus groups of young persons in the design phase of special surveys may provide useful information, and although it is still in the pioneering phase, its use is growing. Different approaches may be used to gain different goals. For instance, Spruyt-Metz (1999) used focus groups to pretest a self-administered questionnaire on health and risk behavior among Dutch high school students aged 12-17. She was mainly interested in question interpretation and the meaning of important concepts and used open interview type of questioning. Cannell, Camburn, Dykema, and Seltzer (1992) used focus groups of American adolescents to test the acceptability of health-related sensitive topics (e.g. cigarette smoking). They presented subjects with potential questions, and stimulated group discussion by giving specific probes on the understanding of the question, how one would react, whether or not one would answer it, or answer it truthfully.

A rather unorthodox, but fruitful application of focus group techniques was employed by Watson, Denny, Adair, Ameratunga, Clark, Crengle, Dixon, Fa'asisila, Merry, Robinson and Sporle (2001) in New Zealand, who used *post pilot* focus groups to evaluate the usability of Multimedia CASI techniques. Following completion of a questionnaire, students aged 12-18 participated in structured focus groups. Each group consisted of six to ten students of the same gender and took about 40 minutes. Open-ended questions were used to stimulate the

discussion. Topics were available time, use of headphones, use of computer, but also question difficulty and emotional burden of the questions. The focus groups revealed two important themes. First of all, the students were very positive about the multi media computer interface and especially the audio component. In the eyes of the respondents the computer made everything easier. The second perceived advantage was privacy. Students appreciated the computer, but also emphasized how important it was that nobody else could read the screen.

Focus Groups with Children and Adolescents

Compared to general adult focus groups, focus groups for children and adolescents appear to be more structured and more centered around specific tasks. Whether this is inherent for groups with children and young adolescents, or whether this is the result of the specific topics in the studies cited above, is unclear. The researchers do not describe in detail if and how the focus groups were adapted to the younger respondents. However, general publications about interviewing children (e.g., Wilson and Powell, 2001; Delfos, 2000) emphasize the importance of a well-designed protocol for open interview situations and the extreme importance of explaining clearly what is expected of the child. This is also stressed by Morgan, Gibbs, Maxwell, and Britten (2002), who wrote one of the first methodological articles about focus groups with children.

Although children and adolescents are acquainted with group discussions, as in most countries, class begins with a short group of pupils and teacher, they will not know what a *focus* group is and what the rules of a focus group are. Therefore it should be made very clear to them what is expected and also that a focus group is not school or a test situation. Also, during the focus group itself, the participants sometimes need to be reminded of the rules. For instance, Morgan, Gibbs, Maxwell, and Britten (2002) wrote simple rules on a flipchart in the beginning and left these on display during the whole session. Examples of these rules were: everyone gets a chance to speak, speak one at a time; you do not have to put up your hand to talk (this is not school). Of course, explaining the rules is important when conducting focus groups with adults too. But young respondents are still developing the cognitive and social skills for meta-communication (see also section 2) and compared to focus groups for adults, the moderator has to pay more attention to meta-communication.

In general, many issues and good practices for focus groups with adults are common to conducting focus groups with children and with adults; it is a question of translating these

good practices to the needs of certain age groups (Morgan, Gibbs, Maxwell and Britten, 2002). Through the setting and the explicit verbal and nonverbal behavior of the moderator, the researcher has to create a different interaction-stimulating environment for each age group. In the following paragraphs we discuss optimal focus group settings for different age categories, emphasizing the special needs of each group. We will not discuss the general rules for conducting good (adult) focus groups; for a thorough introduction we refer to Morgan (1997) and Stewart and Shamdasani (1990); for a quick overview see Cheng, Choi, Easley and Jackson (1997) and ASA (1997). However, as certain topics, such as group size, and homogeneity, are recurrent methodological issues in focus group set-ups for developing questionnaires (Bishoping and Dykema, 1999), we will explicitly comment on these topics.

1. Group Size

Young children need more attention than older children, and a general rule is the younger the participants, the smaller the group. For children in early middle childhood (ages 7 to 10) a group size of about five is optimal. To increase motivation and keep the attention of these young children, one moderator should constantly be tuned in on motivating the children and keeping the conversation going. A second moderator will be necessary for general practical assistance in running a group of young children (see also, Greig and Taylor, 1999; Morgan, Gibbs, Maxwell, and Britten, 2002). More grown-ups in the room will disrupt the balance of power in the group, and it is advisable to have note takers in a separate room, and videotape the entire session for nonverbal cues and interactions (Annon, 1994).

In late middle childhood and early adolescence (ages 10 to16) group sizes may range from 5 to 8 (Scott, Brynin, and Smith, 1995). A second moderator will no longer be needed for practical child-care issues and may be replaced by a note taker or observer. In late adolescence (16 to 18) group size may increase to 8 to 10 participants, only slightly less than in adult groups (cf. Bishoping and Dykema, 1999).

2. Group Homogeneity

Group composition is an important consideration in focus groups. Homogeneity in age with small age bands (e.g., ages 7 and 8, 9 and 10) is recommended (Morgan, Gibbs, Maxwell, and

Britten, 2002). In early adolescence this is crucial, as the eldest will in general look down on the youngest, who has just left primary school. Here, one should be extremely strict regarding age homogeneity, and separate the 12 and 13 year olds from the older children (cf. Scott, Brynin, and Smith, 1995).

Whether or not groups should be homogenous with respect to gender is age-dependent. Before the age of 10, gender homogeneity is not necessary, but in late middle childhood and early adolescence it is advisable (Greig and Taylor, 1999; Scott, Brynin, and Smith, 1995). In late adolescence much depends on the topic of the study and on culture. For instance, Spruyt-Metz (1999) varied the composition of focus groups of Dutch adolescents. She used both all girl and all boy groups, but also added mixed gender groups to stimulate discussion. According to Spruyt-Metz (1999), having opposite sex members in the group may reduce 'acting out behavior' and make the group more task-oriented. Only for the adolescents of Turkish and Moroccan origin were all groups gender homogenous, because of cultural taboos on discussing many of the topics in the protocol with members of the opposite sex. The findings of Bishoping and Dykema (1999) are helpful in deciding on gender homogeneity for focus groups with late adolescents and young adults (16+). They extensively review the importance of socio-psychological factors in focus groups for adults, and conclude that sex segregation has negative effects, especially on disclosure of emotions and personal information, for men, while for women all-female groups enhance their input.

Scott, Brynin, and Smith (1995) note that their focus groups were homogenous in terms of social economic status. But, this could be country specific and dependent on the schooling system and whether or not there are large status differences between schools, as there are in the United Kingdom.

For all age groups it is advised to avoid having close friends, or even classmates in one group, as this may have consequences for group dynamics. It may stimulate concentration lapses in younger children (Morgan, Gibbs, Maxwell, and Britten, 2002), and inhibit open interactions (Scott, Brynin, and Smith, 1995). Especially in adolescence, when peer pressure is heavy (Gray, 2002), one should avoid selecting children from the same peer groups or school classes and preferably mix children from two or more schools.

3. Duration

The younger the child, the shorter the attention span. In early middle childhood (7-10) the attention span is still limited and this has consequences for the scheduling of a session. One should have short periods of discussion (around 20 minutes) alternated with play activities (Delfos, 2000). Morgan, Gibbs, Maxwell, and Britten (2002) used two 20-minute sessions separated by a short refreshment break; they also advise keeping the (tape) recorder running during the breaks to catch relevant remarks.

According to Delfos (2000) children 10 to12 can have longer periods of discussion (30-45 minutes), alternated with refreshment breaks. Scott, Brynin, and Smith (1995), who studied children aged 11 to 16, used focus groups that lasted approximately two hours. Although the attention span of these older children is longer, the moderator should carefully monitor the process and stimulate participation. Group discussion can be alternated with other activities, such as making lists of important points (Morgan, Gibbs, Maxwell, and Britten, 2002), showing pictures, or having children handle survey material (Scott, Brynin, and Smith, 1995). Adolescents can handle discussion periods of one hour, after which a refreshment break is definitely needed. This is as long as most adult focus groups. Still, one has to remember that young adolescents are not adults. They need more time to think, as their mental processing speed is still lower (cf. Kail, 1993).

4. General setting

Notably with the younger children (7-10), the setting should be chosen with careful consideration of the demand characteristics of the room. The moderators should always be on the same eye-level as the children (Annon, 1994; Delfos, 2000). Annon (1994) also notes that when a one-way mirror is used, it should not be on the same level as the children, as it may distract them. In setting the scene, it is also important to pay attention to the power balance. Morgan, Gibbs, Maxwell, and Britten (2002) explicitly chose an informal arrangement, in which all participants sat on soft mats on the floor in the middle of a pleasant light room in a community center. Furthermore, to reduce the hierarchical adult-child relationship, all used first names and all had colorful buttons with their names.

To promote group cohesion with these young children and to clearly communicate that interaction and participation are the goal of the session, group games are advised as warming up. Morgan, Gibbs, Maxwell, and Britten (2002) used a ball game to introduce the group members to each other; a ball was thrown to a group member who had to state his/her name,

favorite food, animal, etc., and then throw the ball to another participant. This is also very useful to assess the cognitive and verbal development of the children and to tune into the child's language (Cares, 1999).

Similarly with children in late middle childhood (10-12), the setting should be chosen with consideration of the demand characteristics of the room, and the moderators should be on the same level as the children. However, one should avoid treating children this age as little ones, as they feel quite superior to the younger children in primary school. Warm-ups and informal introductions remain extremely important and age related games play an important role in this. When moderators and children draw special name labels together, this helps to get acquainted and to reduce the authority imbalance (Hill, Laybourn, and Borland, 1996). Nonverbal communication is an important part of controlling the group process and at regular times and after each sub topic, the moderator has to structure the session by summarizing and asking for additions from the children (Delfos, 2000).

For adolescents it is extremely important that the setting itself has no relationship at all with school or youth centers. It should be new and neutral territory for all, so none of the adolescents is in a power advantage. Especially for the younger adolescents (12-16), careful monitoring of the group process is recommended, and shy adolescents should be encouraged. One way to do this is alternating the verbal discussions with other tasks. For instance, let each one individually write down what he or she thinks is important. The moderator can ask the more quiet group members what they have written and so reduce dominations of the group by the more boisterous ones. Compared to adult focus groups, more time should be dedicated to warming-up and acquainting the members with the rules and goals of a focus group. All focus groups are vulnerable to group pressure and conformity effects, but adolescents are more sensitive to peer pressure than younger children and adults. With adolescents, moderators have to be even more attentive to group processes, and give feed back when necessary. Finally, the moderators should realize that they themselves are *not* young (even if they are 22) and that fashions, music and fads change very quickly (Isacson, 2002, Personal communication). Moderators should never try to be one of the group, as in participant observation, and should never transcend their older adult identity (cf. Morgan, Gibbs, Maxwell, and Britten, 2002).

In-depth or Cognitive Interviews for Testing Questionnaires

Cognitive interviewing in the context of pretesting questionnaires is a form of in-depth interviewing used to find out what goes on in the head of a respondent when answering questions. The 'cognitive interview' in questionnaire testing should not be confused with the 'cognitive interview' in the context of law and child-witness literature. Although both procedures have the name in common, they are different. The 'cognitive interview' of a child witness is a special structured interview taking the respondent step-by-step back to the event, and explicitly designed to get more reliable reports on past events (e.g., Memon, Holley, Wark, Bull, and Koehnken, 1996; Memon and Koehnken, 1992). To pretest questionnaires thoroughly, in-depth interviews, also called cognitive interviews or verbal reporting, are used to investigate the total question-answer process in-depth and to discover sources of confusion and misunderstanding. This method is widely used as a pretest method to investigate the understanding of questions by *adults*, and has proven to be successful in identifying potential problems in questions and in suggesting solutions for these problems (Campanelli, 1997; Presser and Blair, 1994; Willis, Schechter, and Whitaker, 1999).

Potentially, cognitive pretesting of questions could also be a successful method with children and adolescents. It relies heavily on think-aloud procedures, which come very naturally to children. Young children often talk aloud in a non-communicative manner during play or when performing tasks. According to the Russian developmental psychologists Vygotsky (Gray, 2002) this is a natural and necessary phase in the acquisition and internalization of language and verbal thought. Furthermore, think-aloud procedures are often used as an educational tool in primary and secondary schools, especially in teaching mathematics (Kraemer, 2002; Lynn, 2002). Strangely enough, one of the first studies using cognitive testing procedures with young respondents (age 10-21) reported that think-aloud procedures were problematic and that most teenage respondents lacked the ability or the motivation to spontaneously articulate their thought processes (Stussman, Willis, and Allen, 1993). Blair (2000) also reports problems when using think-aloud protocols with young children (6-11). However, both studies gave standard think-aloud instructions for adults, and the procedures were not adapted for younger respondents. Stussman, Willis, and Allen (1993) suggest that traditional cognitive interviewing techniques need to be modified for the young, with more attention to non-verbal communications, and more probes. In addition, Blair (2000) comments that more introduction and explanation is likely to be necessary for children to be good respondents.

Think-aloud procedures with young respondents can work well, as Hess, Rothgeb, and Zukerberg (1998, see also Zukerberg and Hess, 1996) showed. They conducted cognitive interviews with adolescents aged 12-17 to evaluate question understanding, task difficulty, and question sensitivity for the youth questionnaire in the US Survey of Program Dynamics. The researchers developed a detailed protocol beforehand that included probing questions. They report that during the interviews they found a greater need to probe than they typically do during cognitive interviews with adult respondents. This corroborates the conjecture of Stussman, Willis, and Allen (1993) that the young need more extensive probing.

Levine and Huberman (2002) also successfully used think-aloud techniques to test questions on background information from the U.S. National Assessment of Educational Progress questionnaires with children aged 9 and 13-14. Levine and Huberman (2002) developed a detailed protocol with special probes for the cognitive interviews, and interviewers were trained to use them. The young respondents were given a special instruction and explanation of the procedure. Each think-aloud was preceded by having the respondent read the specific question aloud. This facilitated the detection of language and comprehension problems and served as a warm-up for the think-aloud. During the think-aloud the young respondents were continuously encouraged in a neutral manner and probes were used frequently.

Unique in the Levine and Huberman study is that validating information was available based on responses by parents and teachers, which enables comparison of revised questions with original questions. It is encouraging that Levine and Huberman (2002) showed that revised questions had a lower error rate.

Cognitive Test Interviews with Children and Adolescents

Using cognitive interviews for pretesting of children's questionnaires is possible and can result in worthwhile information, provided that the procedures are adapted to the special needs of children and adolescents. In the following paragraphs we discuss necessary adaptations to the general set-up and protocol for in-depth interviews with adults. To accommodate different age groups, adaptations have to be made to all phases: arrival, introduction, start of the interview, interview, and ending (cf. Snijkers, 2002)

1. Arrival

In early and late middle childhood (7-12) special attention has to be paid to this stage. The child will be accompanied be a parent, caretaker or teacher, and both child and caretaker have to be welcomed and introduced to the interviewer, and time has to be taken to make the young child feel at ease. With children, the arrival stage includes many aspects of the introductory stage too. Confidentiality aspects and background information (why is the study done, etc) have to be explained briefly to both parent and child. Therefore, part of the general procedures that, with adult respondents, are discussed in the introduction of the interview, are now introduced at the arrival stage when the parent or caretaker is still there (e.g., explain video-tape, ask permission to record the session (both parent and child should give permission). In early adolescence more often than not a caretaker will still accompany a child and as consequence the arrival will take more time. With older adolescents, the situation more resembles the usual situation with adults. The arrival takes less time, with confidentiality and consent discussed during the introduction. However, in many countries, consent of a parent or caretaker is needed even for older adolescents (16-18), and should be obtained before the session.

2. Introduction

For a successful cognitive laboratory interview the introduction is crucial. In general one has to take more time to explain what the rules are and what is expected than with adults. The importance of this is illustrated by Presser, Blair, Mack, Ryan, and Van Dyne (1993), who asked youngsters pre-interview questions on what a survey was. They found that neither younger (6-8), nor older children (9-11) had a clear idea what a survey was and what the goals and rules of a survey were. More explanation of question asking and answering is needed with children than with adults.

3. Starting the Interview

Because the situation is completely new, the procedures have to be explained carefully. The interviewer has to give clear examples and practice the required tasks before the interview

starts. For instance, one can rehearse practice think-aloud tasks using simple age-related examples (e.g., a simple arithmetic task, solving a simple puzzle, sort objects, etc). Extra time should be reserved for explanation and practice-exercises, as part of a short training-phase before the real interview starts.

4. The Interview Itself

In general, the same rules of thumb for duration are valid as for focus groups. However the estimates given for focus groups are the maximum possible. Because of the lively nature and potential for interaction, focus groups are in general more relaxed and demand less of the concentration span than an individual in-depth interview. Especially with the youngest age group, one has to watch the child carefully and react to drops in attention.

Different interviewing techniques for different age groups are advised. Think-aloud is very natural for young children (7-10), who often still read aloud. Levine and Huberman (2002) explicitly asked 9 and 13 year olds to start with reading the question aloud. Not only did this stimulate them to think aloud, it also provided clues for further probing. For instance, when a child could not read or pronounce a word correctly, this could indicate a comprehension problem.

During the think-aloud the interviewer has to be continuously alert, reinforce the child and start up the process if the child stops for a moment (ask: why do you stop, if tired/not concentrating, suggest a short break, etc). Both Hess, Rothgeb, and Zukerberg (1998) and Stussman, Willis, and Allen (1993) recommend that the interviewer probes more frequently than with adults, and it is advisable to prepare a probing protocol and train interviewers to use frequent probes (Levine and Huberman, 2002).

In all cases it is very important to make sure that the child feels completely at ease. Although thinking-aloud is quite natural for young children, they will *not* perform well when they feel uncomfortable or watched. Young children can be very open in a situation they trust, but become completely shy and introverted when they find themselves in an unknown situation (Scott, 1997). In some cases it is therefore better to have a parent or caretaker present at the interview. Only when a young child feels comfortable, will he/she perform well.

Paraphrasing is a technique that should not be used with younger respondents. Especially in young middle childhood (7-10) paraphrasing a question will not work, since children this age tend to repeat a question literally.

Late adolescents (16-18) may feel very embarrassed when asked to do a think-aloud. But, paraphrasing combined with direct probes (e.g., what does this word mean, what do you think it means) may give good results in this age group. It is important for adolescents that the interviewer reinforces them and reassures that this is not a school test and that not the adolescent but the questionnaire is evaluated! Adolescents often experience 'fear of failure', and may be unsure about themselves and their performance. Reassurance and frequent reinforcement is far more important for this group than in testing adults (cf. Hess, Rothgeb, and Zukerberg (1998).

Auxiliary Methods: Observation and Debriefing

Observation

Monitoring of standardized interviews and self-administered questionnaire sessions is a relatively quick method that can provide useful additional information during field-tests and pilot studies. Coding schedules developed for interviewing adults (e.g., Lessler and Forsyth, 1996; Fowler and Cannnell, 1996; Oksenberg, Cannell, and Kalton, 1991) are mainly for *verbal* behavior, for example 'interviewer reads verbatim', 'interviewer deviates slightly', 'respondent interrupts', 'respondent asks clarification', etc. Coding schedules for children should have more emphasis on *nonverbal* behavior, since children, and especially younger children in middle childhood will have more motor (movement) behavior. An example is the study of Presser, Blair, Mack, Ryan, and Van Dyne (1993), who developed and tested three interview protocols to measure daily food intake for children aged 6 to 11. They videotaped all test sessions, and applied an extensive coding schedule with specific nonverbal codes for the child (e.g., head shaking, nodding, smiling) added to the standard verbal coding schedule of Oksenberg, Cannelll, and Kalton (1991) for interviewer behavior. Presser, Blair, Mack, Ryan, and Van Dyne (1993) found that in the younger group, the interviewer deviated twice as much from verbatim reading of the questions as in the older group and used more probes, indicating more problems in the question-answer process. They also found that younger children smiled about three times as much as older children. This could indicate that young children will smile, or laugh to hide that they do not understand a question. However, the fact that it is possible to reliably code overt children's nonverbal behavior, does not necessarily

mean that the interpretation is clear. In the field of child interviews, there is little work on the interpretation of coded behaviors and more research and development is necessary. The new emerging field of usability testing with children (Hanna, Risden, and Alexander, 1997) is facing similar problems, forcing researchers to acquire more methodological knowledge about children as subjects (Markopoulos and Bekker, 2002).

There are few examples of systematic observation of children during pilot testing of self-administered questionnaires. Researchers mainly suffice with noting down the time it takes to fill in a test or questionnaire, to acquire data to improve planning the major fieldwork. An exception is the work of Helweg-Larsen and Larsen (2001, 2002), who observed both standard mainstream and special education students, aged 15-16, while they completed a pilot version of a Danish health survey. The special education students who had learning and problems took longer and read at such a slow rate that they lost grasp of what had just been asked in the text. It became apparent that students in special education, but also a number of mainstream students, experienced literacy problems.

Debriefing

Interviewer and respondent debriefing studies have proved to be useful for studying response errors in survey data (e.g., Campanelli, Martin, and Rothgeb, 1991), and the observations of trained interviewers may provide worthwhile information on difficulties encountered when interviewing children. Until now this promising area has not been explored.

Van Hattum and De Leeuw (1999) who compared computer assisted self administered questionnaires with paper and pencil questionnaires in Dutch primary schools, used a form of teacher debriefing, in which they asked teachers about their experiences, the experiences of their pupils, and problems encountered during data collection. According to the teachers, asking sensitive questions (e.g., about bullying) by computer was less stressful than paper questionnaires for their young pupils (aged 9-12). Teachers also reported problems their pupils had in understanding several questions (e.g., meaning of certain words), but did not report any problems with the computer itself.

There are several examples of the use of respondent debriefing in surveys of the young. Helweg-Larsen and Larsen (2001, 2002) in Denmark, and Watson, Denny, Adair, Ameratunga, Clark, Crengle, Dixon, Fa'asisila, Merry, Robinson, and Sporle (2001) in New

Zealand, added special debriefing questions at the end of computer assisted questionnaires for adolescents. Topics included the computer interface, as well as privacy issues. Hess, Rothgeb, Zukerberg, Richter, LeMinistrel, and Moore (1998) included debriefing questions in a field test of the youth questionnaire of the US Census Survey of Program Dynamics. Like Scott (1997), they used a combination of walkman and self-administered questionnaire, and at their debriefing focused on the reactions to the audiocassette and privacy issues. Based on the debriefing results, the procedures were slightly modified with regard to reduce repetition of the answer categories on tape.

Summary Part 2

Above we discussed various methods for pretesting questionnaires for children and adolescents. For the clarity of this chapter, we discussed each method separately, but this does not mean that in survey practice only one method should be used. In our opinion it is not either-or; the methods discussed in this chapter complement and reinforce each other and should be used in combination. This is clearly illustrated in the study of Presser, Blair, Mack, Ryan, and Van Dyne, 1993, see also Blair, 2000), who used a variety of methods when developing interview protocols for food intake aimed at children aged 6-11. Besides think-aloud pretests, they also compared different interview protocols and videotaped these for behavior coding. The same videotapes were also used as starting point in debriefing interviews. Data from all sources were combined to devise a new interview protocol for food intake. Another good example is the study by Reynes (2002, see also Reynes and Lorant, 2001), who used a combination of pretest methods when adapting the Buss and Perry Aggression Questionnaire to young French children, aged 8-10. Experts were used to check the simplified vocabulary and sentence structure, the questionnaire was then pretested on 8 year olds to make sure that all questions were understood, and in the final phase a pilot study was done on a large sample of 8-10 year olds (N=500) to check psychometric properties such as reliability of the aggression scale. Hess, Rothgeb, Zukerberg, Richter, LeMinistrel, and Moore (1998) used a similar procedure and combined the results of cognitive think-aloud interviews with those of a full field pretest to investigate potential problems in a self-

administered questionnaire of adolescents (12-17) as part of the US Survey of Program Dynamics.

Watson, Denny, Adair, Ameratunga, Clark, Crengle, Dixon, Fa'asisila, Merry, Robinson, and Sporle (2001) and Helweg-Larsen & Larsen (2001, 2002) followed a slightly different procedure when pretesting health surveys for adolescents in New Zealand (12-18) and Denmark (15-16): after having completed the questionnaire in a pilot study, the respondent immediately took part in post-pilot focus groups to investigate their experiences of the survey. Helweg-Larsen and Larsen (2001, 2002) also used systematic observation during the pilot.

Usually cognitive laboratory methods are used in a **pretest**, which is followed by a pilot or field test and the final study, but cognitive laboratory methods can also be useful as **post**-test to gain insight into problems encountered during data collection or data analysis. Questionnaire test methods can be extremely useful after a survey is completed and when unexpected results are found, or in ongoing or longitudinal surveys. The goal of these questionnaire *post-tests* is to identify sources of measurement errors encountered in the data. A prime example is the study of Jakwerth, Stancavage and Reed (1999) who used standardized in-depth interviews to investigate reasons for the high item nonresponse rate reported over the years in achievement test, for eight graders (approximately 13-14 years) of the U.S. National Assessment of Educational Progress.

Although in most research disciplines the instrumentation is checked, the methods vary. For instance, in test development for educational research, an instrumentation phase is always included in which psychometric reliability and validity of the test are estimated on a large sample, while a cognitive pretest of the questionnaire is rarely employed. In survey research, cognitive pretests are being used increasingly, and pave the way for the costly pilot phase. In our opinion a cognitive pretest should always be part of the test design stage. It is very cost-efficient and gives a thorough insight in what may be wrong with questions and test-items and suggests ways to improve them.

Appendix A: Checklist for Questionnaire Development for Children and Adolescents. Age Groups Early Middle Childhood (7-10), Late Middle Childhood (10-12), Early Adolescence (12-16), Late Adolescence (16-18)

Question features	Description	Practical Recommendations
Comprehension and Interpretation of Question		
Readability index for the introduction text expressed by grade level (USA grades)	Number of words in the introduction. Number of sentences in the introduction. Number of syllables in the introduction. Only the introduction counts and not stem of the questions or the response scale. These aspects will be combined in a readability-index expressed in grade level[1]	Rewrite introductions that have a predicted practical grade level higher than the actual grade of the intended group The grade level predicted by this readability test means that about 50% of pupils in this grade level could read the question (official grade level). For practical use the advice is that questions ought to be about two grades less than the official calculated population grade (or two years less than the official reading age).
Readability index for the question expressed by grade level (USA grades)	Number of words in the introduction. Only the stem of the question counts and not the response scale. Number of sentences in the question. Only the stem of the question counts and not the response scale. Number of syllables in the question. Only the stem of the question counts and not the response	Rewrite questions that have a predicted practical grade level higher than the actual grade of the intended group For practical use the advice is that questions ought to be about two grades less than the official

[1]Flesch-Kincaid readability formula (This is a US government Department of Defense standard tests and results in US school grades, see Johnson, 2002: http://www.timetabler.com/reading.html)
Calculate L, the average sentence length (number of words / number of sentences)
Calculate N, the average number of syllables per word (number of syllables/ number of words).
US Grade level $=(L \times 0.39) + (N \times 11.8) - 15.59$. To get the reading age comparable to grade level add 5. So, Reading Age $= (L \times 0.39) + (N \times 11.8) – 10.59$ years

The grade level predicted by this readability test means that about 50% of pupils in this grade level could read the question (official grade level). For practical use the advice is that questions ought to be about two grades less than the official calculated population grade (or two years less than the official reading age).

Question features	Description	Practical Recommendations
	scale. These aspects will be combined in a readability-index expressed in grade level	calculated population grade (or two years less than the official reading age).
Ambiguity of the question	A question should be conceived as ambiguous if the stem of the question can be interpreted in different ways. In that case the meaning of the question is not the same for all respondents. Vague quantifiers as 'sometimes' or 'often' but also sentences like 'doing thins at home very often' should be coded as ambiguous.	Avoid ambiguous questions. Children in middle childhood (7-12) cannot cope with ambiguity at all. Young adolescents can cope better, but still experience major problems (12-16). Still difficult for older adolescents (16+)
Ambiguity of the response scale	Response scales should be conceived as ambiguous if the response categories of the question can be interpreted in different ways. In that case the meaning of the response scale is not the same for all respondents. Example: vague quantifiers as 'sometimes' or 'often'	Avoid ambiguous response scales. Children in middle childhood cannot cope with ambiguity at all. Young adolescents can cope better, but may experience major problems. Still difficult for older adolescents
Depersonalised or indirect question	Most people ….. People my age ….	Children in young middle childhood do not understand this. Elder children may still experience problems continuing to 12-13
Double barreled question	If the question includes more than one question, it should be conceived as double-barreled.	Avoid double barreled questions
Complex construction of the question	A question should be conceived as complex if the question includes complex constructions. Examples are constructions with semicolons, colons, brackets, parenthesis, or subordinate clauses.	Avoid complex constructions Amount of complexity that can be handled is increasing with age. Children in middle childhood very sensitive (7-12), which continues in early adolescence very sensitive to

Question features	Description	Practical Recommendations
		complexity (12-16). Can remain problem 16+
Negatively formulated question	If there is an option of double negative when answering the question it should be conceived as negatively formulated too.	Avoid negative formulated questions. Children in middle childhood (7-12) cannot handle negations. Double negatives should be avoided at all costs
Suggestively phrased question, statements	E.g. " X" is good (agree/disagree)	Early middle childhood (7-10) extremely sensitive to slightest suggestion, Yeah-saying, agreeing to statements

Retrieving Relevant Information from Memory

Question features	Description	Practical Recommendations
Complexity of the question	A question should be conceived complex if the information that is being asked for in the question cannot directly be retrieved from memory. In complex questions the expectation of the researcher is that the respondent all retrieve kinds of information units from memory, combine these information units and compare them. Examples: how often do you go to the library?	Avoid complex questions Memory of children (capacity & constructive processes) not full-grown before 11. Processing speed comparable to adults after early adolescence (around 15-16)
Reference period	Does the question ask for an undefined or defined reference period?	If a reference time is applicable avoid undefined reference periods, use a defined reference period. Young middle childhood (7-10) many difficulties with past, should be 'here & now' questions
Numerical quantity	Does the question ask for a numeric quantity?	Avoid questions that ask for a numeric quantity. Definitely not suited in middle childhood (7-12). Memory & constructive processes not full-grown before

Question features	Description	Practical Recommendations

Judgment: Information Integration and Evaluation

Question features	Description	Practical Recommendations
Sensitivity of question	Is this question too personal for most children? and/or Is this question too threatening for most of children? and/or Do most children rather not answer this question? and/or Do most children find it hard to give an honest answer to this question?	Sensitivity is not an issue in early middle childhood (7-10), but in later middle childhood (10-12) approval seeking becomes an issue, and in adolescence peer pressure increases. Group norms play an important role. Assure privacy in classrooms
Balance of the question	A question should be conceived as balanced if an equal number of positive response options as negative options offered. In other words a symmetric response scale. A question should also be conceived as unbalanced if the stem of the question mentions one side of the possible answers. Example: Do you agree with the following statement, while offering response options agree versus disagree	Avoid unbalanced questions. May convey suggestions, especially important in early middle childhood
Position in the questionnaire		Randomize question order within blocks. Order blocks so that early questions direct to information to increase cognitive accessibility later

Reporting: Comprehension of Response Options and Selection of Response Option

Question features	Description	Practical Recommendations
Number of response categories	What is the number of response options?	Optimal number of response options early middle childhood (7-10) equals 2-3, later middle childhood and early adolescence 4-5. Graphical (e.g., smileys) more
Offering midpoints	Does the response scale include a neutral	Offer only when it can be

Question features	Description	Practical Recommendations
	midpoint?	clearly labeled (or pictured smiley)
Scale labels	Are the response categories completely, partly or not labeled?	Always offer completely labeled response options (except when using pictorial response scales, e.g., smileys, the pictorials **are** nonverbal labels and verbal labels are not necessary)
Offering don't know filter	Is an explicit Don't know filter offered in the question?	Only offer an explicit *Don't know* options if there are indications for the absence of children's attitudes or knowledge. Especially young children opt out with 'do-not-know'
Offering visual response cards	Are visual response cards offered?	Offer visual response cards and visual stimuli to help memory capacity. Telephone interview not before adolescence. Early adolescence (12-16) simple telephone response categories (3 max)

References

Abramowitz, R., Freedman, J.L., Henry, K., and van Brunschot, M. 1995 "Children's capacity to agree to psychological research: knowledge of risks and benefits and voluntariness" *Ethics & Behavior* 5: 25-48.

Aldridge, M. and Wood, J. 1998, *Interviewing children: A guide for childcare and forensic practitioners*, New York: Wiley.

Amato, P. and Ochiltree, G. 1987. "Interviewing children about their families. A note on data quality." *Journal of Marriage and Family* 49: 669-675.

Annon, J.S. 1994, "Recommended guidelines for interviewing children in case of alleged sexual abuse." *IPT Journal*, Volume 6. Retrieved May 28 2002 at: http://www.ipt-forensics.com/journal/volume6/j6_3_2.htm

ASA 1997. "What are focus groups?" *ASA series what is a survey*. Alexandria: American Statistical Association, Section on Survey Research Methods.

Brainerd, C. and Ornstein, P.A. 1991, "Children's memory for witnessed events: The developmental backdrop." in J. Doris, (ed.). *The suggestibility of children's recollections* (pp.10-20). Washington, DC: American Psychological Association.

Beebe, T.J., Harrison, P.A., McRae, J.A. jr., Anderson, R.E., and Fulkerson, J.A. 1998, "An evaluation of computer-assisted self-interviews in a school setting." *Public Opinion Quarterly* 62: 623-632.

Belli, B. Personal Communication, December 11, 2002.

Benson, J. and Hocevar, D. 1985, "The impact of item phrasing on the validity of attitude scales for elementary school children." *Journal of Educational Measurement* 22: 231-240.

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., and Sudman, S. (eds.) 1991, *Measurement errors in surveys*, New York: Wiley.

Bishoping, K., and Dykema, J. 1999, "Towards a social psychological programme for improving focus group methods of developing questionnaires." *Journal of Official Statistics* 15: 495-516.

Blair, J. (2000), "Assessing Protocols for child interviews." in A.A. Stone, J.S.Turkkan, C.A.Bachrach, J.B. Jobe., H.S. Kurtzman, and V.S.Cain, (eds.), *The science of self-*

report; Implications for research and practice (pp.161-174), Mahwah N.J.: Lawrence Erlbaum.

Borgers, N., de Leeuw, E., Hox, J. 1999, "Surveying children: Cognitive development and response quality in questionnaire research." in A. Christianson (ed.), *Official Statistics in a Changing World* (pp.133-140), Stockholm: SCB.

Borgers, N., de Leeuw, E., and Hox, J. 2000, "Children as respondents in survey research: Cognitive development and response quality." *Bulletin de Methodologie Sociologique (BMS)* 66: 60-75.

Borgers, N. and Hox, J. 2001, "Item non-response in questionnaires with children." *Journal of Official Statistics* 17: 321-335.

Borgers, N. and Hox, J. 2002, "Reliability of responses in questionnaire research with children." Submitted.

Borgers, N., Hox, J., and Sikkel, D. 2003, "Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral midpoint." *Quality and Quantity* (accepted).

Borgers, N. 2003, *Questioning children's responses*, Amsterdam: TT-Publikaties.

Brainerd, C. and Ornstein, P.A. 1991, "Children's memory for witnessed events: The developmental backdrop." in J. Doris (ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp.10-20), Washington, DC: American Psychological Association.

Campanelli, P. 1997, "Testing survey questions: New directions in cognitive interviewing." *Bulletin de Methodologie Sociologique (BMS)* 55: 5-17.

Campanelli, P., Martin, E., and Rothgeb, J. 1991, "The use of respondent and interviewing debriefing studies as a way to study response error in survey data." *The Statistician* 40: 253-264.

Campanelli, P. (2008). "Testing survey questions". In: E.D. de Leeuw, J. J. Hox, and D.A. Dillman (eds). International Handbook of Survey Methodology (pp176-200). New York: Lawrence Eralbaum associates (LEA), Taylor and Francis group.

Cannell, C.F., Camburn, D., Dykema, J., and Seltzer, S. 1992, *Applied research on the design and conduct of surveys of adolescent health behaviors and characteristics,* Ann Arbor: ISR, University of Michigan.

Cares 1999, "Preschool children. Notes from a clinical response to child abuses." CARES Northwest Conference 1999. Retrieved May 28 2002, at http://www.geocities.com/wellesley/8819.html

Ceci, S.J., and Bruck, M. 1993, "The suggestibility of the child witness: A historical review and synthesis." *Psychological Bulletin* 113: 403-439.

Cheng, L., Choi, C., Easley, D., and Jackson, J. 1997, "Focus groups." *Indiana University Continuous learning project.* Retrieved May 28 2002 at http://universe.indiana.edu/clp/rf/focusg.htm

Cole, C.B., and Loftus, E.F. 1987, "The memory of children." in S.J.Ceci, M.P.Toglia, and D.F.Ross, (eds.). *Children's eyewitness memory,* New York: Springer.

Conger, J.J. and Galambos, N.L. 1996, *Adolescence and youth: psychological development in a changing wold* (5th ed.), New York: Longman.

Converse, J. M., and Presser, S. 1986, *Survey Questions. Handcrafting the standardized questionnaire* (Vol. 63), Newbury Park: Sage.

Cynamon, M.L. and Kulka, R.A. 2001, "Collecting data from children and adolescents." Introduction, in M.L.Cynamon, and R.A.Kulka, (eds.), *Seventh Conference on Health Survey Research Methods,* Hyattsville: Department of Health and Human Services.

Czaja R. and Blair, J. 1996, *Designing surveys: A guide to decisions and procedures,* Thousand Oaks: Pine Forge Press.

Delfos, M.F. 2000, *Luister je wel naar mij? Gespreksvoering met kinderen tussen vier en twaalf jaar* [In Dutch: Are you really listening to me? Interviewing children between 4 and 12], Amsterdam: SWP, WESP-publicatiereeks.

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

Dillman, D.A. 2000, *Mail and Internet Surveys. The Tailored Design Method* (2nd ed.), New York: Wiley.

Eisenhower, D., Mathiowetz, N.A., and Morganstein, D. 1991, "Recall error: Sources and bias reduction techniques." in P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz, and S. Sudman, (eds.), *Measurement errors in surveys* (pp.127-144), New York: Wiley.

Esomar 1999, "Guideline on interviewing children and young people." Retrieved November 8 2003. http://esomar.nl/guidelines/interviewing_children_99.htm

Esposito J.L. and Rothgeb, J.M. 1997, "Evaluating survey data: making the transition from pretesting to quality assessment." in L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), *Survey measurement and process quality*, New York: Wiley.

Flavell, J.H. 1985, *Cognitive development* (2nd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Flavell, J.H., Miller, P.H., and Miller, S.A. 1993, *Cognitive development* (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Foddy, W. 1996, *Constructing Questions for Interviews and Questionnaires; Theory and Practice in Social Research,* Cambridge: Cambridge University Press.

Fowler, F.J. 1995, *Improving survey questions* (vol.38), Thousand Oaks: Sage.

Fowler, F.J. and Cannell, C.F. 1996, "Using behavioral coding to identify cognitive problems with survey questions." in N. Schwarz, and S. Sudman, (eds.), *Answering Questions* (pp.15-36), San Francisco: Jossey-Bass.

Gelman, R. and Baillargeon, R. 1983, "A review of some Piagetian concepts." in P.H. Mussen, (ed.), *Handbook of child psychology* (vol.3, 4th ed.), New York: Wiley.

Goodwin, C.J. 2002, *Research in psychology; Methods and design* (chapter 2), New York: Wiley.

Gray, P. 2002, *Psychology*, New York: Worth Publishers.

Greig, A., and Taylor, J. 1999, *Doing research with children,* London: Sage.

Groves, R.M. 1989, *Survey errors and survey costs*, New York: Wiley.

Hallfors, D., Khatapoush, S., Kadushin, C., Watson, K., and Saxe, L. 2000, "A comparison of paper vs. computer-assisted self interview for school alcohol, tobacco, and other drug surveys." *Evaluation and Program Planning* 23: 149-155.

Hanna, L., Risden, K., and Alexander, K.J. 1997, "Guidelines for usability testing with children." *Interaction* 9-14.

van Hattum, M.J.C., and de Leeuw, E.D. 1999, "A disk by mail survey of pupils in primary schools; data quality and logistics." *Journal of Official Statistics* 15: 413-430.

Helweg-Larsen, K. and Larsen, H.B. 2001, "Experiences from a pilot study: The potential for conducting a national questionnaire study on the well-being of adolescent school children, with a particular focus on the sexual encounters with adults." Paper presented to the Danish Ministry of Social Affairs, prior to the national surveys on youth, well being with a focus on CSA.

Helweg-Lasen, K., and Larsen, H.B. 2000, "A nation-wide survey conducted among 15-16 year old Danish school children that included questions on sensitive topics." Paper presented at the International conference on improving surveys (ICIS-2002), August 2002, Copenhagen.

Hershey M., and Hill, D.B. 1976, "Positional response set in pre-adult socialization surveys." *Social Science Quarterley* 56: 707-714.

Hess, J., Rothgeb, J.M., and Zukerberg, A. 1998, "Developing the survey of program dynamics survey instruments." SM 98/07 US Bureau of the Census. Retrieved November 1 2002, http://www.census.gov/srd

Hess, J., Rothgeb, J.M., Zukerberg, A., Richter, K., LeMinistrel, S., and Moore, K. 1998, "Teens talk: Are adolescents willing and able to answer survey questions?" Paper presented at the annual meeting of the American Association of Public Opinion Quarterly in St.Louis. Retrieved November1 2002 at
http://www.amstat.org/sections/srms/onlineproceedings

Hill, M., Laybourn, A., and Borland, M. 1996, "Engaging with children about their emotions and well-being: methodological considerations." *Children and Society* 10: 129-144.

Holaday, B., and Turner-Henson, A. 1989, "Response effects in surveys with school-age children." *Nursing Research* (methodology corner) 38: 248-250.

Isacson, M. 2002, "Personal Communication." June 2002

Jakwerth, P.M., Stancavage, F.B., and Reed, E.D. 1999, "An investigation of why students do not respond to questions." NAEP Validity Studies, American Institute for research, Palo Alto, CA, USA.

Johnson, M.K. and Foley, M.A. 1984, "Differentiating fact from fantasy. The reliability of children's memory." *Journal of Social Issues* 40: 33-50.

Kail, R. 1993, "The role of global mechanism in developmental change in speed of processing." in M.L. Howe, and R. Pasnak, (eds.), *Emerging themes in cognitive development*, New York: Springer.

Kail, R. 1990, *The development of memory in children* (3rd ed.), New York: W.H. Freeman.

Knäuper, B., Belli, R.F., Hill, D.H., and Herzog, A.R. 1997, "Question difficulty and respondents' cognitive ability: The effect on data quality." *Journal of Official Statistics* 13: 181-199.

Kohlberg, L., and Puka, B. 1994, *Kohlberg's original study of moral development. The development of modes of moral thinking and choice in the years 10 to 16* (vol.3), New York: Garland.

Kraemer, J-M. 2002, "Evaluer pour mieux comprendre les enfants et ameliorer sa pratique" (In French: Better understanding of children and improvement of educational daily practice through evaluation), Arnhem: National Center for Evaluation of Education, Cito.

Krosnick, J.A. and Fabrigar, L.R. 1997, "Designing rating scales for effective measurement in surveys." in Lyberg, L., Biemer, P, Collins, M., de Leeuw, E., Dippo, C., Schwarz, N.,

and Trewin, D. (eds.), *Survey measurement and process quality* (pp.141-164), New York: Wiley.

La Greca, A.M. 1990, "Issues and perspectives on the child assessment process." in A.M. La Greca, (ed.), *Through the eyes of the child. Obtaining self-reports from children and adolescents* (pp.3-17)*,* Boston: Allyn and Bacon.

Law and Contemporary Problems 2002, (vol.65, #1) "Special Issue On Children Learning to learn." 2002, <u>Reading</u>. Retrieved, 6 June 2002, http://snow.utoronto.ca/Learn2/mod6/reading3.html

de Leeuw, E.D. and Otter, M. E. 1995, "The reliability of children's responses to questionnaire items; question effects in children questionnaire data." in J.J. Hox, B.F. van der Meulen, J.M.A.M. Janssens, J.J.F. ter Laak, and L.W.C. Tavecchio, (eds.)., *Advances in family research* (pp.251-258), Amsterdam: Thesis Publishers.

Lessler, J.T. and Forsyth, B.H. 1996, "A coding scheme for appraising questionnaires." in N. Schwarz and S. Sudman, (eds.), *Answering Questions* (pp.259-291)*,* San Francisco: Jossey-Bass.

Levine, R. and Huberman, M. 2002, "What types of survey items can elicit valid responses from fourth and eighth grade students." Paper presented at the QDET 2002-conference, Charleston, November 2002. See also http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=200119 and http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=200206, both retrieved September 26 2002.

Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N., and Trewin, D. (eds.) 1997, *Survey measurement and process quality,* New York: Wiley.

Lynn, 2002, "Personal Communication." September 2002

Maccoby, E.E. and Maccoby, N. 1954, "The interview: A tool of social science." in G. Lindzey, (ed.), *Handbook of social psychology*, volume I, theory and method (pp 449-487), Cambridge, Mass: Addison-Wesley.

Markopoulos, P. and Bekker, M. 2002, "How to compare usability testing methods with children participants." in M. Bekker, P. Markopoulos, and M. Kersten-Tsikalkina, (eds.), *Interaction design and children,* Proceedings of the international workshop interaction design and children (pp.153-158), August 2002, Eindhoven/Maastricht: Shaker Publishing.

Market Research Society 2003, "Guidelines for research among children and young people." Retrieved January 17 2003. http://www.mrs.org.uk/fr-code.htm

Memon, A., Holley, A., Wark, L., Bull, R., and Koehnken, G. 1996, "Reducing suggestibility in child witness interviews." *Applied Cognitive Psychology* 10: 503-518.

Memon, A., and Koehnken, G. 1992, "Helping witnesses to remember more: The cognitive interview." *Expert Evidence: The International Digest of Human Behavior, Science, and Law,* 1(2): 39-48.

Morgan, M., Gibbs, S., Maxwell, K., and Britten, N. 2002, "Hearing children's voices: methodological issues in conducting focus groups with children aged 7-11 years", *Qualitative Research* 2: 5-20.

Morgan, D.L. 1997, *Focus groups as qualitative research* (revised edition), Thousand Oaks: Sage.

Nelson, N. 1976, "Comprehension of spoken language by normal children as a function of speaking rate, sentence difficulty, and listeners' age and sex." *Child Development* 47: 299-303.

Oksenberg, L., Cannell, C.F., and Kalton, G. 1991, "New strategies for pretesting survey questions." *Journal of Official Statistics* 7: 349-365.

Otter, M. E. 1993, *Leesvaardigheid, leesonderwijs en buitenschools lezen. Instrumentatie en effecten* [Reading ability, education in reading, and reading outside the school setting: Development of measurement instruments and a study of effect size], Amsterdam: SCO, University of Amsterdam.

Piaget, J. 1929, *The child's conception of the world,* London: Routledge & Kegan Paul (also *Introduction to the child's conception of the world,* New York: Harcourt. [original text: La causalité physique chez l'enfant; la representation de la monde chez l'enfant, Paris, 1927].

Presser, S. and Blair, J. 1994, "Survey Pretesting: Do different methods produce different results?" *Sociological Methodology* 24: 73-104.

Presser, S., Blair, J., Mack, K., Ryan, C.M., and van Dyne, M.A. 1993, *Final report on the university of Maryland-USDA cooperative agreement to improve reporting for children in the continuing survey of food intakes by individuals*. Survey Research Center of the University of Maryland.

Ralph, A., Williams, C., and Campisi, A. 1997, "Measuring peer interactions using the adolescent social interaction profile." *Journal of Applied Developmental Psychology* 18(1): 71-86.

Reynes, E . 2002, "Personal Communication", October 2002.

Reynes, E. and Lorant, J. 2001, "Do comparative martial arts attract aggressive children?" *Perceptual and Motor Skills* 94: 21-25.

Riley, A., Rebok, G., Forrest, C., Robertson, J., Green, B., and Starfield, B. 2001, "Young children's reports of their health: A cognitive testing study." in M.L. Cynamon, and R.A. Kulka (eds.), "Seventh Conference on Health Survey Research Methods." Hyattsville: Department of Health and Human Services.

Saywitz, K. 1987, "Children's testimony: age related patterns of memory errors." in S.J. Ceci., M. Toglia, and D. Ross (eds.), *Children's eyewitness memory* (pp.36-52), New York: Springer.

Schwarz, N., Park, D., Knäuper, B., and Sudman, S. (eds.) 1999, *Cognition, aging, and self-reports*, Philadelphia, 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, and D. Trewin, (eds.)., *Survey measurement and process quality (*pp.331-350), New York: Wiley.

Scott, J., Brynin, M., and Smith, R. 1995, "Interviewing children in the British household panel survey." in J.J. Hox, B.F. van der Meulen, J.M.A.M. Janssens, J.J.F. ter Laak, and L.W.C. Tavecchio (eds.), *Advances in family research (*pp 259-266), Amsterdam: Thesis.

Selman, R. 1980, *The growth of interpersonal understanding,* New York: Academic Press.

Snijkers, G.J.M.E. 2002, *Cognitive laboratory experiences: On pretesting, computerised questionnaires and data quality* (chap.4), Utrecht: University of Utrecht [published Ph.D. thesis].

Spruyt-Metz, D. 1999, *Adolescence, affect, and health* (chap.5), London: Psychology Press.

Stewart, D.W. and Shamdasani, P.N. 1990, *Focus groups; Theory and Practice.* London: Sage.

Stussman, B.J., Willis, G.B., and Allen, K.F. 1993, "Collecting information from teenagers: experiences from the cognitive lab." *Survey Research Method Section (SRMS) of the 1993 ASA conference proceedings.* Retrieved November 1 2002 http://www.amstat.org/sections/srms/onlineproceedings

Sudman, S., Bradburn, N.M., and Schwarz, N. 1996, *Thinking about answers,* San Francisco: Jossey-Bass.

Svallfors, S. 2002, "Personal Communication." December 2002.

Swanson, H.L. 1999, "What develops in working memory? A life span perspective." *Developmental Psychology* 35: 986-1000.

Tanur, J.M. (ed.) 1994, *Questions about questions; Inquiries into the cognitive bases of surveys*, New York: Russel Sage Foundation.

Tourangeau, R. and Rasinski, K.A. 1988, "Cognitive processes underlying context effects in attitude measurement.", *Psychological Bulletin* 103: 299-314.

Turner, C.F., Ku, L., Rogers, S.M., Lindberg, L.D., Plaeck, J.H., and Sonenstein, F.L. 1998, "Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology.", *Science* 280: 867-873.

University of Essex 2002, "Legal and ethical issues in interviewing children." Retrieved January 17 2003. http://www.qualidata.essex.ac.uk/creatingdata/guidelineschildren.asp

Vaillancourt, P.M. 1973, "Stability of children's survey responses.", *Public Opinion Quarterly* 37: 373-387.

Watson, P., Denny, S., Adair, V., Ameratunga, S., Clark, T., Crengle, S., Dixon, R., Fa'asisila, M., Merry, S., Robinson, E., and Sporle, A. 2001, "Adolescents' perceptions of a health survey using multimedia computer-assisted self-administered interviews.", *Australian and New Zealand Journal of Public Health* 25: 6.

Willis, G.B., Schechter, S., and Whitaker, K. 1999, "A comparison of cognitive interviewing, expert review, and behavior coding: What do the tell us?" Survey Research Method Section (SRMS) of the 1999 ASA conference proceedings. Retrieved November 1 2002 http://www.amstat.org/sections/srms/onlineproceedings

Wilson, J. C., and Powell, M. 2001, *A guide to interviewing children: essential skills for counselors, police lawyers and social worker*, New York: Routledge.

Zill, N. 2001, "Advantages and limitations of using children and adolescents as survey respondents." in M.L.Cynamon, and R.A. Kulka, (eds.), *Seventh Conference on Health Survey Research Methods*, Hyattsville: Department of Health and Human Services.

Zukerberg, A.L. and Hess, J. 1996, "Uncovering adolescent perceptions: Experiences conducting cognitive interviews with adolescents." SM 96/01 US Bureau of the Census (also presented at the 1997 AAPOR-conference Orfolk, VA). Retrieved November 1 2002, http://www.census.gov/srd