Understanding and Improving the Validity of Self-Report of Parenting

Sarah K. Morsbach[1] **and Ronald J. Prinz**[1,2]

Published online: 25 April 2006

Although researchers using parental self-report data have questioned its validity (Holden, 2001) and called for more work in this area (Krevans & Gibbs, 1996; Locke & Prinz, 2002), methodological concerns regarding self-report about parenting practices have not been addressed adequately. The susceptibility of parental report to systematic distortions suggests a need for research on ways to improve self-report. This paper attempts to synthesize research findings from other fields (e.g., survey methodology) about ways to improve the validity of self-report, and to discuss the implications for self-report of parenting behaviors. Methods for improving self-report are presented for each of the 5 major tasks in responding to a question: (1) understanding the question, (2) recalling relevant behavior, (3) inference and estimation, (4) mapping the answer onto the response format, and (5) "editing" the answer for reasons of social desirability. Self-administered interviewing, audio-computer-assisted interviewing, pretesting, conversational interviewing, and the decompositional item may be among the best candidates for use in parental self-report. Recommendations are offered with respect to strategies that might prove useful in improving parental assessment of parenting, and to research efforts to evaluate the utility and potential costs of these strategies.

KEY WORDS: parenting; self-report; children; assessment.

Research on parenting relies heavily on parental self-report for assessing attitudes, behaviors, and feelings. Studies comparing self-report with alternate forms of evidence, such as records and behavioral observation, have established that measurement error and conscious bias occur in self-reports (Baker & Brandon, 1990; Belli, Traugott, Young, & McGonagle, 1999). Although researchers have questioned the validity of parental self-report (Holden, 2001; Perepletchikova, & Kazdin, 2004) and called for more work in this area (Krevans & Gibbs, 1996; Locke & Prinz, 2002), methodological concerns regarding parental self-report about parenting constructs have not been addressed adequately. In addition, few attempts have been made to systematically test methods for improving parental self-report.

The aim of the current paper is to discuss the state of parental self-report of parenting behaviors and review methods for improving the accuracy of those reports. Parenting practices have been linked with a number of specific child outcomes (Holden, 2001; Macoby & Martin, 1983), including maladaptive outcomes such as aggression and conduct problems (Reid, Patterson, & Snyder, 2002; Stoff, Breiling, & Maser, 1997). Valid measurement of parenting practices thus has important implications for the study of clinical child and family psychology.

Parenting constructs have particular elements that are vulnerable to the distortions inherent in self-report. First, parents are often asked to make estimates of potentially high-frequency behaviors (e.g., conversation, yelling) over long periods of time (e.g., a month or more), a cognitively difficult task. Increases in cognitive burden have been associated with decreases in accuracy in self-report, as respondents are more likely to use less precise estimation strategies when responding to items

[1] University of South Carolina, Columbia, South Carolina.

[2] Address all correspondence to Ronald J. Prinz, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208; e-mail: prinz@sc.edu.

requiring more cognitive effort (Tourangeau, Rips, & Rasinski, 2000). Second, there is uncertainty as to the degree of consensus in the general population about the definition and interpretation of certain terms related to parenting, such as time-out (Clayman & Wissow, 2004). Third, many parenting items can be considered highly sensitive in nature. Tourangeau et al. (2000), for example, identified three dimensions of sensitivity: social desirability, intrusiveness, and risk of disclosure to third parties. Parenting items often have response choices that are associated with a certain degree of social desirability. In a sample of American parents, Morsbach and Prinz (2004) found high levels of consistency in perceptions of the socially desirability of different parenting behaviors. For example, 98.2% of parents indicated that praising one child is generally considered to be a good way to parent. Intrusiveness pertains to an item's inherent offensiveness. Some inquiries about parenting may be viewed as offensive, even by parents who feel confident that their accurate responses place them at the socially desirable pole. For instance, a parent may perceive questions regarding the manner in which she disciplines her child to be intrusive, even if she believes that her discipline strategies are not generally perceived as undesirable. The third dimension of sensitivity, risk of disclosure to third parties, may also apply when assessing parenting behaviors. Some parents may fear legal reprisals for truthfully reporting about their discipline practices. Thus, social desirability, intrusiveness, and risk of disclosure are all relevant in considering the sensitivity, and hence validity, of parental self-report. Therein lies the challenge: a large body of research has demonstrated a strong tendency for socially desirable responding in the face of sensitive questions (Bradburn, 1983; Schaeffer, 2000; Schwarz, 1999a,b), and parental self-report is not immune from this issue.

Since the early 1980s, researchers from a variety of disciplines have been studying cognitive processes underlying self-report, as well as methods to improve its accuracy. This research has been labeled the Cognitive Aspects of Survey Research Movement (CASM; Jabine, Straf, Tanur, & Tourangeau, 1984). This paper attempts to synthesize some of the CASM findings and to discuss implications for parental self-report. Specific emphasis will be placed on self-report of parenting practices, rather than on attitudes or other parenting constructs, and on the questionnaire format, although some of the findings reviewed can also be applied to interview and other formats. The first section of the paper provides an overview of the current state of parental self-report in the parenting field, discussing common measures as well as concordance between parental self-report and other measures of parental behavior, such as behavioral observation. The second section will discuss interventions that have been found to improve the accuracy of self-report in other areas that may be able to be applied to self-report of parenting behaviors.

THE CURRENT STATE OF PARENTAL SELF-REPORT

In evaluating the validity of parental self-report, one difficulty is the lack of a gold standard to which self-report can be compared. In other fields, such as that of substance abuse, self-report can be compared with physical measures of recent drug use behavior, via blood tests or urinalysis. Similarly, self-reports of abortion or voting can be compared with actual records of those activities. With parenting, however, the task of corroborating self-report is more complex.

One strategy to increase validity is the use of either multiple informants or methods to measure the same construct. On the basis of findings from research on self-report of addiction, Del Boca and Doll (2000) suggested three benefits to the use of multiple alternate sources. First, respondents tend to report more accurately when they know or believe that corroborating evidence (e.g., other-report, observation) will be collected. Second, given the strengths and weaknesses of each data type, different sources of data can be triangulated to provide more accurate information than would any single source. Third, differences in the nature or magnitude of discrepancies between two data sources across time or treatment group can be assessed. Such differences may reflect the influence of methodological artifacts.

Systematic observation as an alternate source has been recommended as a strategy to improve the validity of parental constructs (Fagot, 1992). Although generally considered to be a sound recommendation, the use of observation comes with its own set of challenges and limitations. In addition to its associated costs, observation generally cannot provide a measure of many important parenting behaviors, such as harsh physical discipline (Peterson, Tremblay, Ewigman, & Popkey, 2002) or leaving a child at home without supervision (Shelton, Frick, & Wootton, 1996). In addition, children's reactivity to

observation tends to increase with age (Ollendick & Hersen, 1984; Shelton et al., 1996), which makes it more of a challenge to observe parents interacting with older children. Despite these limitations, observation provides a unique and important source of data on parenting, best utilized in conjunction with self-report or other methods (Holden, 2001). In addition to systematic observation, another alternative is report by others, which can be used in conjunction with self-report. Researchers have collected reports of parenting behaviors from children (e.g., Simons, Whitbeck, Conger, & Wu, 1991) and from other parents or caregivers (e.g., Lovejoy, Weis, O'Hare, & Rubin, 1999).

Accuracy of Parental Self-Report

Because it is difficult to verify the accuracy of parental self-report, few studies have directly measured the extent to which parents are accurate self-reporters. One exception is in the area of parental report of child vaccination histories. Several studies have documented a tendency for parents to considerably overreport the number of vaccinations their child has received (Goldstein, Kviz, & Daum, 1993; Willis, Brittingham, Lee, Tourangeau, & Ching, 1999). Willis et al. (1999) asked parents whether their children had received various vaccinations, and found little relationship between parental reports and the child's vaccination status, as verified by medical records. False positive rates ranged from 67 to 100% for different vaccines, suggesting a tendency by parents to report that their child had received a vaccine regardless of whether or not this was the case. When asked a global question about whether or not their children's vaccination status was up to date, 83% of parents whose children were not up to date answered in the affirmative. A study by Goldstein et al. (1993) yielded more accurate responses by parents to the global question, yet one-third of parents still answered the question incorrectly. More accurate estimates in the Goldstein et al. (1993) study may be due to the fact that their population included only parents of younger children (3–5 years) whereas Willis et al. (1999) included parents of children up to age 13. Parents of younger children have fewer vaccinations to recall than that by the parents of older children.

Willis et al. (1999) hypothesized three factors contributing to low rates of accuracy in parental report of child vaccinations: (1) parents may never have encoded the relevant information in the first place, (2) they may not have been able to recall the information when asked, and (3) they may have edited their responses because of a reluctance to admit that their child was not fully vaccinated (i.e., social desirability concerns). Findings from a study in which parents were asked about their child's vaccination status immediately after the child received vaccinations suggested lack of encoding to be a prominent reason for low rates of accuracy (Willis et al., 1999).

The implications of these studies for self-reports of parenting are difficult to determine given the lack of direct studies of the accuracy of self-reports about parenting behaviors per se. It should not be assumed that parents respond with such low rates of accuracy to simpler questions (e.g., "How often do you attend meetings at your child's school?") However, the set of studies on parental report of vaccinations empirically demonstrates that parents are not always accurate reporters, that under certain circumstances accuracy ratings can be quite low, and that memory difficulties and socially desirable responding are potential forces in inaccurate responding.

Measurement of Parental Self-Report

Before considering potential strategies to improve self-report of parenting, it is relevant to examine the current status of measurement in this area. Numerous measures have been developed to assess a wide variety of parenting domains (Holden, 2001). An important element to consider is the internal consistency of these measures, as well as their concordance with other sources of data on parenting.

Eight measures of self-report of parenting practices were selected on the basis of two comprehensive reviews (Holden, 2001; Locke & Prinz, 2002), as well as the availability and quantity of associated psychometric data. These eight parenting report measures are described later and are summarized in Table I with respect to internal consistency and concordance between sources. It is important to note that the measures presented in this paper represent a sample of the measures of parenting practices that are currently available and in clinical use. Readers are referred to Holden (2001) and Locke and Prinz (2002) for a more comprehensive review of parenting practice measures.

The Parent Behavior Inventory (PBI; Lovejoy et al., 1999) was developed for use with children in preschool or early elementary school. Twenty items assess broad areas of parenting that have been identified as problematic in clinic-referred families:

Table I. Psychometric Properties of a Sample of Self-Report Measures of Parenting

| Instrument | Sample population | | Scale(s) (internal consistency) | Observation | Source of psychometric data |
	Parents of	n			
APQ—Parent Global Format	Children aged 6–13 years	160			Shelton et al. (1996)
			Involvement (.80)		
			Positive parenting (.80)		
			Poor monitoring/supervision (.67)		
			Inconsistent discipline (.67)		
			Corporal punishment (.46)		
CAP	Review article summarizing 9 studies		Physical child abuse: (.74–.94)		Milner (1994)
CRPR	Primary school children	239		Nurturance (.11)	Dekovic et al. (1991)
			Nurturance (.74)		
			Restrictiveness (.83)	Restrictiveness (.43*)	
			Authoritative (.71)	Authoritative (.25*)	
			Authoriatrian (.65)	Authoritarian (.44*)	
LYQ—Parent Form	Children aged 10–18 years	192			Jacob et al. (2000)
			Bad friendships: (.82, .77)[a]		
			Shared recreational activities (.85, .77)		
			Child/parent involvement (.74, .67)		
			Child involvement with family (.53, .56)		
			Curfew (.67, .70)		
			Persistence of discipline (.46, .64)		
			Parent–child relationship (.84, .87)		
			Positive parenting (.78, .75)		
			Supervision (.45, .60)		
			Parental time spent with child (.83, .85)		
MCBS	Children aged 2–14 yaers	396	Total (.84)		Perepletchikova and Kazdin (2004)
PBI[b]	Preschool-aged children	IC: 107, PR: 46, OB: 50			Lovejoy et al. (1999)
			Supportive/engaged (.83)	Supportive/engaged (play: .11, task: .50*, task)	
			Hostile/coercive (.81)	Hostile/coercive (play: .30*, task: .54*)	
PAQ-R—Parent self-report version	Children in Grades pre-K through 5	:	Authoritarian (.76, .75, .72)[c]		Reitman et al. (2002)
Sample 1		87[d]	Permissive (.74, .74, .73)		
Sample 2		102[e]	Authoritative (.77, .74, .73)		
Sample 3		171[f]			
PS	Mothers of children aged 18–48 months	IC: 168, OB: 15			Arnold et al. (1993)
			Laxness (.83)	Laxness (.61*)	
			Overreactivity (.82)	Overreactivity (.65*)	
			Verbosity (.63)	Verbosity (.53*)	
			Total (.84)	Total (.73*)	

Note. APQ: Alabama Parenting Questionnaire; CAP: Child Abuse Potential Inventory; CRPR: Child Rearing Practice Report; LYQ: Loeber Youth Questionnaire; MCBS: Management of Children's Behavior Scale; PBI: Parent Behavior Inventory; PAQ-R: Parental Authority Questionnaire—Revised; PS: Parenting Scale; IC: internal consistency; OB: observation; PR: parent report.

[a] The first score is for maternal report and the second score is for paternal report.
[b] Partner report: supportive/engaged, .26 ($p < .05$); hostile/coercive, .42 ($p < .05$).
[c] Coefficient alphas presented for Samples 1, 2, and 3, respectively.
[d] Parents in the sample were predominantly Caucasian with higher SES.
[e] Parents in the sample were predominantly African American with lower SES.
[f] Parents of children aged 3–5 years enrolled in Head Start Program.
*$p < .05$.

support/engagement and hostility/coercion. Also used with the parents of preschool children, the Parenting Scale (PS; Arnold, O'Leary, Wolff, & Acker, 1993) focuses on discipline practices that relate to externalizing behavior problems in young children. Thirty items are scored using a response format with polar anchor points: less adequate parenting at one end and more adequate parenting at the other end.

The Parental Authority Questionnaire (PAQ; Reitman, Rhode, Hupp, & Altobello, 2002) was developed for use with parents of children in grades pre-K through 5. This 30-item measure assesses attitudes and behaviors associated with authoritative, authoritarian, and permissive parenting styles. Used with parents of both elementary and middle school aged children, the Alabama Parenting Questionnaire (APQ; Shelton et al., 1996) was designed to measure aspects of parenting practices related to disruptive behavior problems. The measure comprises 42 items rated on a frequency Likert scale and has five content scales. The Loeber Youth Questionnaire (LYQ; Jacob, Moser, Windle, Loeber, & Stouthamer-Loeber, 2000) assesses parenting practices related to the development of aggressive and antisocial behavior in preadolescent and adolescent youth (aged 10–18 years). Fifty-eight items comprise 10 scales.

Rather than focusing on one particular age group, some measures can be used with children of a wide variety of ages. The Child Abuse Potential (CAP) Inventory (Milner, 1994) has 160 items relating to different types of child maltreatment. Block's (1965) Child Rearing Practice Report (CRPR) measures child-rearing behaviors, attitudes using a Likert-type scale. The questionnaire format consists of 91 items. Finally, the Management of Children's Behavior Scale (MCBS; Perepletchikova & Kazdin, 2004) was developed for parents of children aged 2–14 years. This 38-item measure tests parenting practices related to the development of child conduct problems.

For the measures described previously and reviewed in Table I, internal consistency estimates for the various subscales range from .45 to .94. Reliability coefficients of .80 or higher are generally considered to be desirable (Holden, 2001). Although approximately two-thirds of the coefficients reported in the studies reviewed were lower than .80, the majority of the scales produced internal consistency estimates of greater than .70, indicating that the internal consistency of most measures is close to the desired level. Taken together, these estimates suggest the need for improvements in internal consistency of parenting measures.

Also notable is the fact that concordance between parental self-report and observation of parenting was measured and reported for only three of the measures.

Concordance between parental self-report and observation can be especially difficult to interpret, as many observational and self-report items do not measure the same behaviors (Lovejoy et al., 1999), and self-report typically encompasses a wider reference period (e.g., 1 month, 3 months) than that which constitutes an observation period (e.g., 30 min). Two studies found significant correlations between self-report and observation despite different reference periods (Arnold et al., 1993; Dekovic, Janssens, & Gerris, 1991). Another indicator of validity of parental self-report is the extent to which self-report is concordant with reports from other individuals, such as children or spouses. Several researchers have reported significant correlations between parent and child report of parenting constructs for adolescent children (e.g., Krevans & Gibbs, 1996; Simons et al., 1991). However, when comparing parental self-report with younger children's reports of parenting, researchers have not found evidence of convergent validity (e.g., Sessa, Avenevoli, Steinberg, & Morris, 2001; Shelton et al., 1996). In a rare study comparing mother and father appraisals of maternal parenting behaviors, Lovejoy et al. (1999) found significant correlations for supportive/engaged and hostile/coercive constructs from the PBI ($r = .26$ and $r = .42$, respectively).

In summary, several studies supported moderate concordance between parent and child reports, particularly in the area of discipline, with estimates ranging from .23 to .37. One study of measures reviewed in this paper provided evidence of convergent validity between reports from two parents. Comparisons between observation and parental self-report tended to yield higher estimates, ranging from .15 to .63. Overall, there is sufficient evidence to show some degree of convergent validity between parental self-report and other methods of measurement. The next important step is to identify the conditions that are necessary to produce more reliable and more valid parental self-report. Examination of findings from studies of self-report methodology in other fields and disciplines may be a good starting place.

METHODS FOR IMPROVING SELF-REPORT

Methods for improving self-report are presented for each of the five tasks identified by Schwarz and Oyserman (2001) in responding to a question: (1) understanding the question, (2) recalling relevant behavior, (3) inference and estimation, (4) mapping the answer onto the response format, and (5) "editing" the answer for reasons of social desirability. The strategies for improving self-report stem from other research applications besides parenting, such as substance abuse and public health. The extent to which each strategy can be successfully applied to parenting has not yet been established empirically. The strategies included are those that might be applicable to parenting.

Understanding the Question

Ideally, survey questions should be clear and unambiguous, having the same meaning for each respondent and the researcher. Schwarz (1999a, 1999b) distinguishes between literal and pragmatic meanings of a question. The literal meaning of a question pertains to the semantic understanding of the words and their definitions and is important but not sufficient for a respondent to adequately understand the question. A pragmatic meaning requires the respondent to understand and make inferences about the interviewer's intention. Take, for example, a parent who is asked, "What have you and your child done together today?" The grammar and definitions of terms are seemingly clear—the question is asking what activities the child and parent have engaged in together so far during the current day. Despite the clear literal meaning, the pragmatic meaning requires inferences as to what kinds of activities the investigator is interested in hearing about. Should the parent's response include driving in the car together and sitting in the same room before breakfast, or is the researcher interested primarily in more active pursuits, such as playing a game or going to a friend's house? Thus, giving an appropriate answer to a question requires both a literal and pragmatic understanding of the meaning of a question. In this section, three types of strategies will be discussed that have been shown to help respondents understand both the literal and pragmatic meanings of a question.

Strategies for Writing Clear Questions

The first strategy, or set of strategies, involves creating survey items that are as clear and unambiguous as possible. These strategies will be described briefly in this paper, but are discussed at length in a number of books, edited volumes, and review articles (see Oppenheim, 1992; Schwarz, 1999a,b; Tanur, 1991; Tourangeau et al., 2000). Some of the strategies have empirical support for improving question comprehension (e.g., defining ambiguous terms), whereas others derive more from logic (e.g., avoiding complicated syntax). Typical sources of comprehension errors will be presented, along with the associated strategies for improving survey items. In many cases, the strategy is simply to avoid the error source.

Incorrect, ambiguous, or complex syntax can result in errors of literal (mis)understanding. Questions constructed to minimize such errors should typically avoid multiple embedded clauses, structures with double negatives, ambiguous terms that have not been defined, and overly complex cognitive tasks (Tanur, 1991). Tasks or questions that overtax respondents' working memory will often lead to the use of less precise estimation strategies. If necessary, questions that cover multiple possibilities can be decomposed into multiple questions to simplify the task.

Errors of pragmatic meaning include presupposition, unfamiliarity, and vagueness (Tourangeau et al., 2000). When faced with an item containing a presupposition that does not apply to a respondent, he or she may respond in a number of different ways, such as asking for clarification or responding "I don't know," "no," or "never" (Lessler, Tourangeau, & Salter, 1989). For example, an item that asks parents how often they attend school meetings and meetings for other activities their child is involved in presupposes that the child attends school. For nonschool-aged children, some parents might report "never," others "n/a" and still others might respond on the basis of their attendance at nonschool activities, making the distribution of responses difficult to interpret. Unfamiliar terms can pose a similar problem, leading to a variety of responses.

Finally, vague concepts and vague quantifiers can lead to increased measurement error. In response to a question about the effects of watching violent television programs on children, Belson (1981) found that respondents varied widely in their

interpretation of the term "children," with estimates ranging from 8 years and under, to 20 years or younger. A number of studies (e.g., Moxey & Sanford, 1993; Schaeffer, 1991) have demonstrated that the way in which a respondent interprets a scale consisting of adverbial quantifiers depends on a variety of different factors, such as the typical frequency of the event, individual differences among respondents, and the context (e.g., previous questions asked). When they are used, vague quantifiers should be thought of as specifying a relative position in an ordinal series, rather than equivalent to a numerical distribution. This issue is particularly important in the area of parenting, given the field's use of vague quantifiers in a number of common parenting measures. Research is needed to explore the meaning of such quantifiers, the strategies parents use in choosing a response, and whether findings based on vague quantifiers differ from those based on numerical frequencies.

Cognitive Interview

The cognitive interview, a tool for questionnaire development, has been described as the most tangible result of the CASM movement (Conrad & Blair, 1996). The primary goal of cognitive interviewing is to gain insight into the cognitive processes respondents employ when answering specific survey items, and to use that insight to construct better survey questions (DeMaio & Rothgeb, 1996). In particular, components such as the retrieval strategies used, the sequence of thought processes, and the thought contents are analyzed (Bickart & Felcher, 1996). The term cognitive interviewing has a broad scope, subsuming nine methods for pretesting survey questions: concurrent think-alouds, retrospective think-alouds, focus group discussions, confidence ratings, paraphrasing, sorting, response latency, probes, and memory cues (Tourangeau et al., 2000). These methods are routinely used in major survey centers to aid in the development of survey items, but are used less frequently in psychological research, where there is a tendency to construct questionnaires in a more ad hoc, as-needed manner (Schwarz, 1999a,b).

Once the cognitive interview has been conducted, the next step is to analyze and interpret data from the video- or audio-recorded interview. Researchers have developed a wide variety of coding systems (e.g., Blair, Menon, & Bickart, 1991; Fowler & Cannel, 1996; Willis, 1997), focusing on different

types of codes and constructs. In one behavioral coding system, Fowler and Cannel (1996) code respondent behaviors such as interruptions, clarifications, inadequate answers, and response refusals. Other coding systems focus on strategies used by the respondent to reach a result (Blair et al., 1991) or distinguish between types of problems encountered by respondents in answering specific items (Conrad & Blair, 1996).

In spite of the widespread use of the cognitive interview, only a few studies have looked at its effectiveness in yielding survey questions with more accurate results (e.g., Fowler, 1992; Lessler et al., 1989; Presser & Blair, 1994). These studies have examined the extent to which survey items modified on the basis of findings derived from cognitive interviewing produce more accurate responses than do the original items. Significant differences were found between the modified and original items for a health survey (Fowler, 1992) and a survey including items on diverse topics such as transportation and commercial expenditures (Willis & Schechter, 1997). Results of a study on dental health were inconclusive regarding the role of the cognitive interview in improving respondent accuracy (Lessler et al., 1989).

Although results are inconclusive regarding its empirically demonstrated effectiveness, the cognitive interview has strong face validity, endorsements from survey researchers in many subject areas, and case studies demonstrating its usefulness. For the most part, neither the cognitive interview protocols nor the coding systems currently being used are specific to a particular subject area. Therefore, these already existing strategies could be used to implement and interpret cognitive interviews on parenting survey items. Particularly given the ambiguous terms, long reference periods, and vague quantifiers used in parenting surveys, the field could potentially benefit a great deal from employing cognitive interview methods to pretest and revise parenting survey items. Using a cognitive interview to pretest an instrument's parenting items might also contribute to a better understanding of the processes by which parents self-report about parenting. Greater detail about the cognitive interview approach can be found in Willis (2004).

Flexible Interviewing

Once a question has been pretested, flexible (also called conversational) interviewing techniques

can be used to help respondents interpret questions in the intended manner. These techniques, employed by the interviewer, might include providing a definition to a word or even presenting the question in his or her own words. Flexible interviewing can be contrasted with standardized interviewing, in which interviewers can use only neutral probes to aid respondents who ask for help.

There is currently a debate among survey researchers as to the effectiveness of standard versus flexible interviewing. Proponents of the standardized interview (e.g., Beatty, 1995; Kovar & Royston, 1990) argue that standardization leads to a greater statistical precision and makes it affordable to test larger populations. They suggest that problems in understanding can be best addressed by designing and pretesting better questions. In addition, flexible interviewing could lead to greater variability rather than stability in situations in which interviewers are not sure how to define terms or help respondents apply their specific situation to the item. Others (e.g., Schober & Conrad, 1997; Suchman & Jordan, 1992) argue that standardization suppresses important elements of ordinary conversation that are used to mediate ambiguities of relevance and interpretation. For example, in a conversation, speakers can accommodate future questions based on responses from previous ones, negotiate mismatches in world-view between the respondent and the instrument, and identify and share uncertainties in meaning or difficulties in understanding. Thus, an interview style that is more flexible and more consistent with the kind of interactions we have during everyday conversation could potentially improve accuracy of responding.

In a series of systematic comparisons of response accuracy for conversational and standardized interviews (Conrad & Schober, 2000; Schober & Conrad, 1997), conversational interviewing did not improve accuracy for items that mapped onto a respondent's life circumstance in a straightforward way but yielded significantly higher accuracy for items that required more complex mapping. This increase in accuracy had a cost; the median interview length for the flexible interview was three times longer than that of the standardized interview (12 min vs. 4 min). These findings suggest that flexible interviewing is most appropriate for use in situations where pretesting indicates that complicated mappings are frequent, but may not be worth the increased expense when few or no complicated mappings are expected.

Flexible interviewing can be utilized on a variety of intensity levels, ranging from most to least standardized (Schober & Conrad, 1997). Interviewers could read scripted definitions of terms only, read scripted probes that offered helpful information, or be given license to improvise in helping the respondent grasp the intended meaning. These interventions could occur only when a respondent asks for help, or could include any instance in which the interviewer feels the respondent is struggling. Schober, Conrad, and Fricker (2004) found comprehension to be most accurate when interviewers provided both requested and unrequested clarifications, responding in their own words (i.e., without standardized scripts).

Although more work is needed in this area, evidence suggests that flexible interviewing is a promising approach for decreasing error due to ambiguity in question meaning. The field of parenting contains a number of potentially ambiguous terms (e.g., slap, time-out, reward) and complex mappings (How would a parent with a custody arrangement that differs from week to week respond about the frequency of specific parenting behaviors in a "typical week?"), making it an appropriate candidate for flexible interviewing. However, the longer duration of interviews could increase cost and might necessitate decreases in sample size. Empirical studies measuring if and when this strategy increases accuracy in parental self-report could help researchers determine under which circumstances the positive effects of the strategy outweigh its costs. Flexible interviewing could be used in combination with appropriate wording and pretesting to increase the likelihood that parents will interpret constructs and question items in the manner intended by the investigator.

Recalling Relevant Behavior

After comprehending an item, the next logical step for the responder is to retrieve relevant information from memory. Retrieval is the process by which information stored in long-term memory is brought into an active state (Tourangeau et al., 2000). Both retrieval cues and a series of other strategies designed to aid memory could be applied to parental self-report, and will be discussed in some detail.

Retrieval Cues

Retrieval cues can be used to help respondents access properties of a specific event (e.g., the most recent school meeting attended) or recall as many

events are possible in a given reference period (e.g., all school meetings attended in the last 6 months). For recalling a specific event, recall cues about what happened have been found to be effective in improving memory for that event (Belli et al., 1999; Brewer, 1988; Wagenaar, 1986). For example, if a researcher wanted to know details surrounding a child's most recent temper outburst, retrieval cues could be used to focus a parent's memory on relevant details, such as the location where the outburst took place, the specific behaviors of the child, or the feelings the parent experienced.

Other types of retrieval cues can be used to assist recall of multiple events, rather than one particular event. Means, Nigam, Zarrow, Loftus, and Donaldson (1989) found that having respondents construct a time line containing personal events as well as the event being cued (in this case, visits to a doctor's office) resulted in recall of additional instances of visits they had not previously reported. However, Chu et al. (1992) did not find greater recall of hunting and fishing activities when using a time line with key events. In identifying how many times a parent lost her temper with her child over a the past week, constructing a time line of events that occurred during the week could help the parent better recall relevant incidents.

An intervention that has received considerable attention is that of decomposition, which involves breaking a larger question into its subcomponents or subcategories (Menon & Yorkston, 2000). For example, if the frequency of interest is the number of times a parent yelled at his or her child in the past week, the question can be decomposed to ask about the number of times a parent yelled during various activities (e.g., mealtimes, bathing, playtime) or settings (e.g., at home, in the car, outside). In addition to easing the cognitive load on respondents, decomposition is also thought to counteract recency, vividness, and saliency effects, each of which makes certain events more accessible in memory and can potentially lead to errors (Menon, 1996). Schwarz (1999a,b), however, warns that decomposition can lead to overestimation, as individuals tend to overestimate lower frequency events; moreover, it does not necessarily lead to better recall. Menon (1996) found regularity of the behavior being measured to moderate the effects of decomposition. Her findings suggest that decomposition would be more likely to lead to enhanced accuracy of recall for behaviors a parent engages in irregularly (e.g., giving the child a tangible reward), but

less likely to improve recall for more regular behaviors (e.g., verbal praise).

Other Interventions

Two memory interventions, which do not include retrieval cues, show promise and could be applied to the parenting domain. These include increasing respondent time on task, and the use of the bounded interview for repeated measures. There is considerable evidence to support the notion that response accuracy increases when respondents take more time to formulate their answers (Burton & Blair, 1991; Cannell, Miller, & Oksenberg, 1981; Means, Swan, Jobe, & Esposito, 1994). Means et al. (1994) instructed respondents to use a variety of strategies and found that taking more time tended to result in greater accuracy of reports of cigarette smoking. Burton and Blair (1991) encouraged participants to take more time to respond and found a positive relationship between response time and recall accuracy of students for B grades received on their school report cards, but not for checks written. They hypothesized that increased time is an effective strategy for memories that are more accessible (B grades received) but not for less accessible memories (checks written). This evidence suggests that the accuracy of parental response to survey items might be enhanced simply by slowing the pace of the interview or instructing parents to take more time in responding.

The bounded interview is a method for reducing telescoping effects, which can occur when respondents are uncertain about the timing of an event. Respondents may mistakenly report events that happened before the start of the reference period (forward telescoping) or fail to report events that happened during the reference period (backward telescoping). Neter and Waksberg (1964) asked respondents to report on household expenditures on two occasions. When, during the second interview, respondents were read a list of expenditures they had reported during the first interview (bounded recall), forward telescoping for the second interview was substantially reduced. In addition, Sudman, Finn, and Lannom (1984) found that respondents reported fewer events in the current month if they had first been asked to recall and report the events of the prior month. Although accuracy of report was not verified by the researchers, results suggested a reduction

of forward telescoping. The bounded interview can be used with repeated measures designs to help prevent overreporting due to telescoping effects. In both studies, event descriptions were reported by respondents (e.g., an expenditure for exterminating the home) in addition to event frequencies (e.g., 25 expenditures last month). In parenting, this strategy could potentially work well for parental reports of specific events (e.g., descriptions and count of major temper tantrums over the past month). It is less clear how well this strategy might work for reports of event frequencies alone (e.g., number of temper tantrums in the previous month).

In summary, cues enhancing retrieval for specific events and retrieval for multiple events, as well as other interventions related to memory are of potential utility to the parenting area. Strategies for recall of specific events, such as retrieval cues, have strong empirical support. However, measurement of parenting constructs generally tends to focus on recall of classes of events rather than a specific event, with the exception of applied behavioral analysis. Evidence for the effectiveness of strategies that aid recall of multiple events such as decomposition and the use of time lines is less conclusive. Decomposition does seem to be a helpful strategy under some circumstances, and could be applied to recall of parenting events that tend to be irregular. For other interventions, increasing the amount of time on task and employing bounded recall appear to show the most promise for the field of parenting.

Inference and Estimation

Once a respondent has retrieved relevant events from memory, the next task is to add up, combine, or summarize these events so they can be used to make judgments. Knowledge about how people make these judgments has been researched extensively by researchers in the areas of recognition memory, decision-making, and survey methodology. However, there has been surprisingly little work on testing interventions to promote more accurate inference and estimation. Of greatest utility to parenting researchers may be an understanding of strategies individuals use when forming judgments, those strategies that tend to produce more accurate responses under which circumstances, and ways to encourage parents to utilize those strategies.

Using methods such as concurrent and retrospective think-aloud probes, researchers have identified a number of strategies used by respondents to formulate judgments (Blair & Burton, 1987; Means & Loftus, 1991; Menon, 1996). The first set of strategies is generally used when respondents have retrieved information about specific relevant events. *Recall and count*, or episodic enumeration, consists of remembering and summing all the events to obtain a frequency. A related strategy, *recall and count by domain*, consists of summing events separately by domain and combining those estimates. *Recall and extrapolate*, also called rate estimation, involves using recall of a few events to estimate a rate, and extrapolating that rate over the reference period. A second strategy type, *tally*, is used with information about a frequency for which the respondent can recall an exact tally. For example, most parents could tell you how many children they have without remembering and counting each child. A third set of strategies uses generic information, in cases where generic representations of events rather than individual events are accessible. *Retrieved rate* consists of recalling existing information about a rate (e.g., I take my child to the park twice a week) and applying that rate to the reference period (e.g., In the last month, I must have taken my child to the park about eight times). A related strategy, *recommended rate*, is generally used when a retrieved rate is not available. For example, in Willis et al. (1999) study described earlier, some parents recalled the number of vaccinations their child had received using their belief about the recommended rate for child vaccinations as an anchor. A final set of strategies, based on general impressions, include *guessing*, also referred to as a rough approximation, and *context-influenced estimates*, in which the middle value on the response scale is used as an anchor and adjusted on the basis of a vague impression.

A number of considerations might affect which strategies are used for a specific item, as well as which strategies produce more precise and accurate responses. These considerations include respondent motivation, accessibility of information, amount of time and level of effort required for the strategy, and a number of task conditions (Tourangeau et al., 2000). Consistent with findings discussed earlier, a general rule of thumb states that rate-based estimation tends to be most accurate when events are regular, whereas recall-and-count strategies are most accurate for infrequent and distinctive events (Tourangeau, 2000). Strategies based on general impression tend to produce the least accurate responses in most circumstances.

A limited number of studies have evaluated interventions to promote the use of more precise and accurate strategies, with mixed results (Blair & Burton, 1987; Bless et al., 1996; Burton & Blair, 1991; Menon, 1996). Blair and Burton (1987) hypothesized that asking respondents "how many times," as compared to "how often," would promote more recall-and-count strategies, but found little support for this hypothesis. Asking participants to respond using a percentage rather than a frequency seemed to promote impression-based strategies (Bless et al., 1996), suggesting that asking for percentages could promote the use of less accurate strategy types. When Burton and Blair (1991) attempted to manipulate both time and motivation, telling respondents that the questions were important and instructing them to take more time to respond, they found respondents more likely to use recall-and-count strategies. Finally, work by Menon (1996) indicates that respondents can be encouraged to use recall-and-count strategies when a larger question is decomposed into smaller categories.

It is a challenge to determine which of the interventions (e.g., decomposing items, instructing participants to take more time, increasing motivation, avoiding the use of percentages, and changing item wording) might be best applied to parenting measures. The complexity of circumstances surrounding a particular measure (likely frequency of events for each item, measure length, respondent motivation) make it difficult to predict those interventions that might be successful for any given participant on any given parenting measure. It appears that one solution is for parenting researchers to apply and evaluate these interventions to promote the use of more precise strategies (e.g., recall and count; inference), rather than strategies based on general impressions.

Mapping the Answer onto the Response Format

Once a judgment has been formed, the next task facing a respondent is mapping this judgment onto the response format provided. The two steps of forming a judgment and mapping the judgment are frequently interrelated: the judgment will often affect the selection of a response option, but, as discussed in the previous section, the response options may also affect the formation of a judgment. Although the mapping stage of processing subsumes many topics, two, in particular, will be discussed in detail because of their potential relevance to the measurement of parenting behavior: a comparison of open versus closed response formats and a discussion of response order effects.

Open and Closed Response Formats

When constructing an item to elicit a numerical behavior frequency, the survey designer must choose either an open or closed format. In an open format, the respondent is simply asked to indicate a number to a question such as, "In the past week, how many meals did you eat with your child?" A closed format version of this question would include response options ranging from 0 to some maximum value, in ranges specified by the researcher (e.g., 0–3, 4–7, etc.).

With the open format, the resulting data is more exact (e.g., 5 meals per week vs. 4–7 meals per week), and it is not necessary to truncate the response options at either end. A disadvantage is the tendency for respondents to provide round answers. Tourangeau, Rasinski, Jobe, Smith, and Pratt (1997) found that reported numbers of sexual partners clustered around multiples of 5, particularly for respondents with larger numbers of partners to report. In effect, respondents by rounding are creating their own sets of response categories. Rounding can also introduce systematic biases, as some respondents may consistently round in a certain direction, either knowingly or unknowingly. There may also be group differences in tendencies to round.

Closed-ended items tend to be easier to code and response options can help clarify the meaning of a question. On the same note, the choice of a response scale can systematically affect responses. In a frequently cited study, Schwarz, Hippler, Deutsch, and Strack (1985) asked respondents to indicate the number of hours of TV they watch daily, providing either a response scale emphasizing the high end of the scale (responses ranging from "up to 2.5 hr" to "more than 4.5 hr") or a response scale emphasizing the low end of the scale (responses ranging from "up to 0.5 hr" to "more than 2.5 hr"). For the response scale with lower options, 16.2% reported watching more than 2.5 hr, compared to 38% for the response scale with higher options. The authors suggested that respondents may view the middle value on the response scale as representing the average or typical value and thus assimilate their response to this perceived average. This assimilation effect may be particularly strong for sensitive questions. Using a

similar design, Tourangeau and Smith (1996) found that respondents reported more sexual partners over the past year for a response scale with higher options (3.38 partners) compared to one with lower options (1.43 partners).

Closed and open question formats each come with their own set of advantages and disadvantages. In general, Tourangeau (2004) recommends the use of the open-ended format, particularly for sensitive items where the assimilation effect may be more likely to occur. A review of the parenting measures discussed in an earlier section suggests that most response scales consist of vague quantifiers, in which an open format is generally not applicable.

Response Order Effects

When asking closed-ended questions with unordered response categories, researchers should be aware of the potential for primacy or recency effects. Unordered response options are categorical rather than numerical in nature. For example, parents might be provided with a list of discipline methods and asked to identify their primary method. According to Krosnick and Alwin's (1987) "satisficing model," participants may choose the first acceptable alternative on the list, rather than the alternative that best answers the item. This primacy effect tends to occur with written or self-administered items. When items are read out by an interviewer aloud, however, recency effects are more likely. Because participants often do not have time to evaluate each response as it is read, they are most likely to remember the last items read and choose the most acceptable among them, even if a response read toward the beginning of the list is more appropriate.

Research indicates that primacy effects are most likely to occur for longer lists, defined as those with five or more alternatives (Schuman & Presser, 1981), but are possible for items with as few as two response options (Schwarz, Hippler, & Noelle-Neumann, 1991). On a voting referendum, Handlin (1994) found modest evidence that respondents were statistically more likely to vote "yes" when yes was first on the list, and "no" when no was first on the list.

A number of strategies that might minimize recency and primacy effects were discussed earlier, such as emphasizing the importance of the research or instructing respondents to take more time and consider each answer. Tourangeau (2004) also suggests randomizing the order of response options.

This is more easily accomplished with computerized administration, where response options can be randomized for each respondent, than with paper-and-pencil administration. Overall, research on response mapping offers valuable information to parenting researchers. Given that closed-ended questions introduce potentially biasing response options, open-ended questions are recommended for questions about behavior frequencies. When responses are categorical, randomizing response order has been proposed as a solution for minimizing primacy and recency effects, though this strategy has not yet been adequately researched. Perhaps the most important contribution of this research is an awareness that response options, whether numerical or categorical, are not neutral and must therefore be chosen with caution.

Editing the Answer

Finally, research has demonstrated that editing or censoring a response is most likely to occur for sensitive topics, with individuals who have something to hide, and when sensitive items are interviewer-administered rather than self-administered (Schaeffer, 2000). Social desirability, a tendency to present oneself in a favorable light, can be an important contributor to editing. Although a few studies of manipulations of item format have been shown to improve self-report under some conditions, findings about interview mode are relatively consistent and robust.

Interview Mode

Although the interviewer- or self-administered paper-and-pencil survey tends to predominate parenting research, a number of other methods are regularly used by the national surveys and other academic projects. Tourangeau et al. (2000) have identified 14 modes, each of which will be briefly described. Telephone modes include (1) the conventional telephone survey, in which an interviewer poses questions from a paper survey and marks answers on a paper; (2) computer-assisted telephone interviewing (CATI), in which the interviewer reads questions off a computer screen and enters responses into the computer; (3) touchtone data entry (TDE), where the respondent answers computer-generated items by dialing a designated number into the phone; and (4) voice recognition entry (VRE), where the respondent verbally indicates a response option that a

computer reads and transcribes. Using the postal service, (5) self-administered questionnaires (SAQ) or (6) disk by mail (DBM) modes involve sending the respondent a paper-and-pencil survey or a disk containing a survey. In addition, (7) web-based surveys as well as the (8) prepared data entry mode (PDE), a type of interactive web survey, have become increasingly common. In person paper-and-pencil modes include (9) paper-and-pencil interviewing (PAPI), where an interviewer reads question from the survey and marks answers; (10) self-administered questionnaire (SAQ), where the interviewer administers a survey and is available to aid the respondent if needed; and (11) audio self-administered questionnaire (ASAQ), in which a recorded voice reads survey items to the respondent and he or she independently marks responses. Finally, three computer-assisted in person modes are (12) computer-assisted personal interviewing (CAPI), where an interviewer reads questions from a computer screen and enters responses; (13) computer-assisted self-administered interviewing (CASI), in which a respondent answers items on a computer with an interviewer present; and (14) audio computer-assisted interviewing (ACASI), where respondents use a computer to respond to questions read by a recorded voice.

These modes differ in five main characteristics, some of which have important implications for socially desirable responding. The first characteristic is the method of contacting the respondent. Respondents can be contacted in person, by telephone, by mail, or through e-mail, among other ways. Although it is clear that these contact methods differ in terms of their access to populations (e.g., the population for telephone contact excludes those without telephones), data regarding the effects of different contact types on response accuracy are ambiguous and at times contradictory (Aquilno, 1994; Groves & Kahn, 1979; Johnson, Houghland, & Clayton, 1989). It is not clear that differences in the method of contact clearly affect responding.

The second characteristic involves the medium of the questionnaire (paper vs. electronic). One particularly robust finding is that the electronic medium, specifically the computer, reduces rates of missing data (Aquilno, 1994; Tourangeau & Smith, 1996; Turner, Lessler, & Devore, 1992). This is likely due to the fact that the computer format makes it more difficult for respondents to accidentally skip items. Regarding response content, most evidence indicates that computer-administration increases candid reporting for sensitive, but not nonsensitive items

(Baker, Bradburn, & Johnson, 1995). Computerization may also slow the pace of interviews, resulting in increases in total administration times (Martin, O'Muircheartaigh, & Curtice, 1993). A slower pace can be considered both an advantage, as it may yield more accurate responses (Cannell et al., 1981), or a disadvantage, as increasing administration time may increase costs.

The third characteristic, method of administration, is the characteristic with the most consistent findings: Respondents report more socially undesirable behaviors when measures are self-administered rather than interviewer-administered. This pattern has been demonstrated in the areas of alcohol and drug use (Aquilno, 1994), religious attendance (Presser & Stinson, 1998), sexual behaviors (Tourangeau & Smith, 1996), and abortion (Mosher & Duffer, 1994). Although a large number of studies have demonstrated this effect, two in particular will be described as an illustration of the impact that self-administration can have. Fendrick and Kim (2001) randomly assigned survey respondents to either a telephone interview, face-to-face interview, or a self-administered questionnaire. The sample consisted of more than 3,000 individuals who had previously disclosed cocaine use. Four years later the same individuals were asked about lifetime cocaine use. Interview mode (both contrasts) was a significant predictor of denial of cocaine use. That is, the mode in which the survey was administered significantly affected which individuals disclosed and which individuals denied cocaine use 4 years later. Highest levels of denial were found for the telephone mode, followed by the face-to-face mode, and finally the self-administered mode.

In a second study demonstrating the advantages of self-administration, Tourangeau and Smith (1996) compared two computerized self-administered modes (ACASI and CASI) to a computerized interviewer-administered mode (CAPI) for reporting of number of sexual partners. A well-established trend is that men tend to overreport number of sexual partners whereas women tend to underreport number of sexual partners, with men reporting as many as three times more sexual partners than do women (Smith, 1992). In a closed population, one would expect females and males to have approximately the same number of sexual partners. In their study, Tourangeau and Smith found that, compared to interviewer-administered, the self-administered modes led to reports of significantly higher frequencies of sexual partners

by women and significantly lower frequencies by men.

Overall, studies demonstrate that self-administration of sensitive items tends to reduce socially desirable responding. Even when an interviewer is present during administration, self-administration has been shown to eliminate variation due to interviewer characteristics (Tourangeau et al., 1997), which can be a source of error for sensitive questions.

Two final characteristics differentiating modes of data collection are the mode of responding (verbally vs. written vs. computerized) and the presentation mode of the items (visual or oral). These characteristics combine to create different levels of cognitive demand. For example, a self-administered questionnaire requires a respondent to have literacy skills as well as the ability to follow routing instructions. The level of cognitive difficulty in completing the survey may affect who can participate in the survey, the amount of missing data, and the reliability of the data (Baker et al., 1995; Martin et al., 1993). When comparing written and oral modes, it is vital to keep in mind the tendency for primacy effects with written material and recency effects with auditory material.

Differences found for the five characteristics discussed earlier likely reflect psychological constructs that affect survey responding. Impersonality pertains to individuals' sense that they are disclosing information to another person. Some modes are more likely to promote feelings of impersonality, such as computer-assisted self-administered interview (CASI) and computer-assisted personal interviewing (CAPI), as they tend to lower the salience of the interviewer in the process. The perceived legitimacy and importance of the study is also an important variable, affecting participation rates, nonresponse, and reports of sensitive behavior (Heberlein & Baumgartner, 1978; Yammarino, Skinner, & Childers, 1991). There is evidence that computerization, independent of whether self- or interviewer-administered, can increase reporting of sensitive behavior (Baker et al., 1995), likely because of an increase in perceived legitimacy. Other factors, such as a survey's affiliation, can also increase perceived importance of a study (Yammarino et al., 1991). Another psychological variable, referred to as cognitive burden, is not fixed by mode but tends to depend on issues such as layout and formatting. (See Jenkins & Dillman, 1997, for a discussion of principles in questionnaire design that relate to cognitive burden.) Finally, confidentiality and anonymity are important variables to consider. In asking students to self-report on cheating behaviors, Ong and Weiss (2000) found that 75% acknowledged having cheated before under conditions of anonymity, whereas only 25% acknowledged cheating under conditions of privacy and confidentiality. Thus, anonymity appears to decrease socially desirable responding above and beyond confidentiality. However, confidentiality alone (without anonymity) has been shown to result in reductions in socially desirable responding for sensitive, but not nonsensitive items, as discussed in a quantitative review by Singer, Von Thurn, and Miller (1995).

Overall, it appears that the most consistent findings occur when comparing different mediums and different forms of administration. For sensitive items, both the computerized mode and self-administration appear to reduce socially desirable responding. A number of studies (e.g., Fendrick & Kim, 2001) were able to check the accuracy of self-reports and thus confirm that this reduction in socially desirable responding results in an increase in accuracy. Psychological variables such as perceived legitimacy, confidentiality, anonymity, cognitive burden, and impersonality play a role in determining the extent to which participants respond in a socially desirable manner. Despite all the new technology available, parenting research tends to rely primarily on self- or interviewer-administered paper-and-pencil surveys. The research strongly suggests that self-administered interviewing offers advantages over interviewer administration for sensitive items (e.g., Aquilno, 1994; Presser & Stinson, 1998; Tourangeau & Smith, 1996). In addition, some data suggest some advantages (e.g., item nonresponse, more reporting of sensitive behaviors) involved with computerization, a technology that could be applied to measurement of parenting. Of course, potential drawbacks associated with computerized self-report (e.g., cost, logistical restraints) must also be considered when deciding what type of administration to utilize.

Item Format

Although research on eyewitness memory (Loftus, 1996) and attitude judgments (Rasinski, 1989) has clearly demonstrated that wording and item format matter, comparatively few studies have looked at the impact of item wording on socially desirable responding for behavioral frequency judgments. Wording effects that have been found tend to be small compared to mode effects (Tourangeau,

2004). Four studies that found item format effects, all using different strategies, will be described in chronological order.

Catania, Binson, Canchola, Pollack, and Hauck (1996) compared an enhanced item to a standard item in a survey of sexual behavior. Enhanced items included preambles with supportive statements about behaviors that could be viewed as nonnormative. For example, before a series of items about sexual partners a preamble stated the following, "The number of sexual partners people have had differs a lot from person to person. Some people report having had one sex partner, some two or more partners, and still others report hundreds of partners" (Catania et al., 1996, p. 352). The control item asked respondents the same series of items but did not include the preamble. The enhanced item generated higher levels of report for same-gender sex, extramarital sex, sexual problems, virginity status before age 18 among men, and more sexual partners, but did not significantly affect reports of condom use. The authors concluded that the enhanced item format was effective across a variety of items, but took significantly longer to administer. Accuracy of responses was not measured using alternate sources, and therefore judgments about the effects of the enhanced items on accuracy were not possible.

Using a different strategy, DiFranceisco, McAulife, and Sikkema (1998) contrasted direct and indirect methods of asking about high-risk sexual behaviors. A sample of homosexual men either were asked the number of times in the past 3 months they had engaged in intercourse "without a condom," an item requiring an explicit admission of practicing unsafe sex, or were asked the frequency of intercourse and the estimated percentage of the time they used a condom. Men who received the second question reported 37% more unprotected intercourse than did men who received the first question. The authors concluded that framing items in terms of positive choices reduces socially desirable responding. One possible confound in this study is that the items may have elicited different strategies (e.g., recall-and-count vs. inference), which could have accounted in part for the differential responding. Again, it is assumed but not verified that the increased responding with the second question reflects more honest reporting.

Belli et al. (1999) were able to verify the self-reported behavior, voting. Survey respondents were asked if they had voted in the most recent election.

An experimental question included both memory cues and a response option that allowed respondents to save face ("I usually vote but didn't this time"). The authors found significantly more agreement between reported and actual voting for the experimental item. They concluded that the response saving option in addition to memory cues increased self-report accuracy of voting behavior, though the design did not allow for the effects of each component to be evaluated separately.

In a recent study of parenting behavior, Morsbach and Prinz (2004) piloted an augmented-permissive item format containing preambles presenting frequency and opposing belief statements. The preambles were designed to reduce socially desirable responding by suggesting either end of the scale to be normative and acceptable. For example, an item on volunteering stated the following, "Some parents enjoy volunteering often with special activities that their child is involved in. Other parents wish they could volunteer more, but do not because they have many other responsibilities. How often do you volunteer to help with special activities that your child is involved in?" The authors found the augmented-permissive item to reduce socially desirable responding for positively perceived parenting items on average, as well as some negatively perceived parenting items. The accuracy of self-reported parenting behaviors was not verified.

Thus, item format interventions that have been shown to work include using supportive language, frequency statements, and cognitive rationales, providing a cognitive rationale as a response option, and positive framing. Each of these interventions shows promise at this time but needs to be replicated.

One final strategy that shows promise is the time-use item. If a researcher is interested in a particular event (e.g., did the respondent exercise on a given day) one approach is simply to ask the respondent to list the activities they engaged in that day, rather than asking directly about exercise. As the time-use item does not make exercise, or any other event, salient, it is hypothesized that the respondent feels less pressure to misreport for reasons of social desirability. In a study on religious attendance (Presser & Stinson, 1998), respondents were contacted by telephone on a Monday and asked to list the activities they had engaged in on the previous day. Compared to an item that asked respondents directly if they had attended services the previous day, claims of weekly attendance were reduced by

one-third with the time-use item. The authors concluded that the time-use item increased accuracy of self-report of religious attendance by removing the social desirability factor. Further study of time-use items is warranted given its potential for mitigating socially desirable responding. Such items could be applied to a number of parenting constructs of interest. For example, if a researcher wanted to sample child-centered versus adult-centered activities, asking for a list of activities might yield less socially desirable responses than simply asking for a list of child-focused activities from the previous day.

Thus, attempts to reduce editing of responses have identified a number of promising interventions that could be applied to parenting, including various changes in item format and response options, the time-use item, computerized technology, and self-administration. Findings about computerization and self-administration have received a great deal of attention, and are more robust than are findings about item format. Statistical interventions, such as measuring and partialing out the effects of social desirability (e.g., Jo, 2000) have also been utilized and show promise. Such interventions are complex, and beyond the scope of this review paper. (See Paulhaus, 1991, for a review of measuring and controlling response bias.)

DISCUSSION AND RECOMMENDATIONS FOR FUTURE RESEARCH

Perhaps the greatest contribution of the CASM movement is not any particular questionnaire strategy but rather an increased awareness and understanding of how multiple factors aside from actual events can contribute meaningfully to self-report estimates of behaviors. In understanding a question, variables such as grammar, syntax, interviewer involvement, and vocabulary can affect the manner in which a respondent interprets, and therefore responds to, a question. Response options as well can provide information to the respondent, contributing to his or her understanding of an item's pragmatic meaning. The effects of these various elements are often difficult to predict without pretesting.

In recalling relevant behavior, the types of retrieval cues included in the item, number and type of retrieval attempts made by the respondent, specificity of the question, reference period length, order of response options, and time taken to formulate a response are all important contributors to what a respondent will recall. In addition, respondents often tend to have difficulty placing events within a particular reference period, and may make telescoping errors.

It is clear that many strategies can be used by a respondent to formulate a response, and that these strategies have differential effectiveness depending on variables such as the length of the reference period, the specificity of the item, the regularity of the behavior, and respondent motivation. Elements such as the time taken by respondents, item wording, examiner instructions, and response options will affect the strategy used by the respondent. When mapping the answer onto the response format, the response format type (closed vs. open), the response options chosen by the experimenter, the order in which response options are presented, and the mode of presentation (visual vs. oral) can contribute to the option chosen by the respondent.

Finally, differences in item wording and interview mode often influence the extent to which a respondent is willing to reveal accurate information in response to sensitive questions. In comparing interview modes, particularly important are the medium of the questionnaire (paper vs. electronic) and the type of administration (self-administered vs. interviewer-administered).

In addition to presenting factors that affect self-report, this paper has reviewed a variety of interventions that may improve its validity (see Table II for a list of these interventions). Although all are applicable to parental self-report in some sense, these interventions vary in the ways in which each can be applied to parenting, the degree to which each is empirically supported, the typical magnitude of effects, and the conditions under which each should be employed. However, the authors are aware of only one study (Morsbach & Prinz, 2004) that has evaluated one of these various interventions in the context of parental self-report. In this sense, much work remains to be done in order to better understand how these interventions might affect the ways in which parents respond to items about their own parenting. Some of these interventions stand out as having better potential to improve parental self-report than others, and will be discussed later.

Perhaps most striking is the contrast between interviewer and self-administration for sensitive items. Numerous studies have demonstrated that self-administration can lead to more accurate responding for sensitive items. Currently, in the field of parenting, both interviewer-administered

Table II Interventions to Improve the Validity of Self-Report: Organized by Five Tasks in Responding to a Question

Intervention
Step 1: Understanding the question
Various strategies for writing clear items
Cognitive interview prior (pretesting)
Conversational interviewing
Step 2: Recalling relevant behavior
Retrieval cues
Increasing respondent time on task
Bounded interview (for repeated measure designs)
Step 3: Inference and estimation
Encouraging a specific estimation strategy or set of strategies
Step 4: Mapping the answer onto the response format
Using open-ended items
Randomize response order for nonordered categories
Step 5: Editing the answer for reasons of social desirability
Self-administration (for sensitive questions)
Computer administration
Take measures to assure confidentiality
Make responses anonymous
Use supportive language and positive framing
Time-use item

and self-administered formats are being used with some frequency. Having an interviewer read items has the potential advantages of addressing literacy concerns and decreasing the amount of missing data. Popular in use with large national surveys, the audio-computer-assisted interviewing (ACASI) mode, in which items are presented orally and visually for respondents in a self-administered format, may offer advantages from both perspectives. Evaluating the effects of this technology in the context of parenting, in relation to its costs, could contribute substantially to the validity of parental self-report of sensitive items.

The cognitive interview, or pretest, is another strategy that could benefit the field in two distinct ways. First, using various techniques and probes, such as asking respondents to think aloud, could help illustrate the processes parents utilize to respond to various types of items. Second, pretesting items can provide valuable information about the ways in which respondents are interpreting specific terms and items. There is some evidence that revisions to items based on pretest findings can result in improved accuracy of self-report. Many protocols and coding systems are available to help parenting researchers implement the cognitive interview and interpret resulting data.

Conversational interviewing shows promise as a strategy for helping parents define terms and interpret items in the manner intended by the experimenter. Interviewers could help parents understand and interpret parenting terms (e.g., time-out, praise) similar to the way in which these terms are interpreted by the experimenter. More work in this area is needed to determine whether improvements in accuracy are worth the expense of increasing the interview duration.

Finally, the decompositional question is a potential method for improving respondent interpretation, memory, and estimation for irregular items. One difficulty in applying this strategy to parenting is that, given the variety in parenting styles and strategies, it is difficult to predict uniformly which parenting behaviors can be considered irregular. Examples of parenting behaviors that may fit into the category of having "irregular" frequencies for most parents are spanking, attending school-related events such as assemblies or Parent Night, or not knowing the whereabouts of a younger child.

Overall, self-administered interviewing, the audio-computer-assisted interviewing (ACASI) mode, pretesting, conversational interviewing, and the decompositional item may be among the best candidates for use in parental self-report. Particularly important will be to not just use but also evaluate the effectiveness and potential costs of these interventions in the parenting context.

In reviewing the current state of parenting measurement, a number of strengths were noted. First, a wide variety of parenting domains have been defined and are being studied, such as various components of discipline, warmth, competency, involvement, and supervision. Second, in order to measure these domains, researchers have developed and are using a number of different methods, including observation, other-report, and self-report. Within each method, multiple measures are available. In some cases, researchers are using multiple methods to measure the same construct and reporting the associated convergence between the methods, which allows for continued evaluation of the validity of specific domains of parental self-report using various instruments.

At the same time, a review of psychometric properties of a sample of measures suggests a need for improvement in internal consistency and convergent validity. Approximately two-thirds of the internal consistency estimates are below the desirable level of .80. When reported, significant correspon-

dence between methods had a median range of approximately .2–.4 for self-report and other-report, and .3–.5 for self-report and observation. On the basis of the current state of parenting and the research contributions from the CASM movement, the following recommendations are offered for future research on self-report of parenting behaviors:

1. Parenting researchers have a number of measures of parenting practices from which to choose. In making a decision, it is important to consider the availability and quality of psychometric data. In cases where it is necessary or desirable to create a new questionnaire, psychometric properties should be evaluated and reported to inform future research.

2. The use of multiple methods and measures to assess a construct is recommended, and is already being utilized by some researchers. In these cases, it is helpful to calculate and report estimates of correspondence. A meta-analysis of correspondence ratings for various methods could be quite informative.

3. A new generation of methodological studies is needed to look at conditions and factors that enhance validity of parental self-report. A good starting point is interventions that have proven particularly effective in other fields and can be applied to parenting, including but not limited to self-administered interviewing, the audio-computer-assisted interviewing (ACASI) mode, pretesting, conversational interviewing, and the decompositional item.

4. Given the frequent use of vague quantifiers (e.g., sometimes, never) in our parenting measures, there is a need to further explore the meaning of these quantifiers. How do respondents select a response option, and how is this process different from choosing from a list of numerical frequencies?

5. An issue particularly important to evaluation of parenting interventions, not specifically addressed by the CASM movement, is that of sensitivity to change in measurement. In addition to identifying measures that are sensitive to behavior changes, it would be helpful to determine the properties of a measure that increase its sensitivity to change.

REFERENCES

Aquilno, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: A field experiment. *The Public Opinion Quarterly*, *58*, 210–240.

Arnold, D. S., O'Leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: A measure of dysfunctional parenting in discipline situations. *Psychological Assessment*, *5*, 137–144.

Baker, R. P., Bradburn, N. M., & Johnson, R. (1995). Computer-assisted personal interviewing: An experimental evaluation of data quality and survey costs. *Journal of Official Statistics*, *11*, 415–434.

Baker, T. B., & Brandon, T. H. (1990). Validity of self-reports in basic research. *Behavioral Assessment*, *12*, 33–51.

Beatty, P. (1995). Understanding the standardized/nonstandized interviewing controversy. *Journal of Official Statistics*, *11*, 147–160.

Belli, R. F., Traugott, M. W., Young, M., & McGonagle, K. A. (1999). Reducing vote overreporting in surveys: Social desirability, memory failure, and source monitoring. *Public Opinion Quarterly*, *63*, 90–108.

Belson, W. A. (1981). *The design and understanding of survey questions*. Aldershot, UK: Gower.

Bickart, B., & Felcher, M. (1996). Expanding and enhancing the use of verbal protocols in survey research. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining the cognitive and communicative processes in survey research* (pp. 115–142). San Francisco: Jossey-Bass.

Blair, E. A., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research*, *14*, 280–288.

Blair, J., Menon, G., & Bickart, B. (1991). Measurement effects in self vs. proxy responses: An information-processing perspective. In P. B. Biemer, R. B. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 145–166). New York: Wiley.

Bless, H., Clore, G. L., Schwarz, N., Golisano, V., Rabe, C., & Wolk, M. (1996). Mood and the use of scripts: Does being in a happy mood really lead to mindlessness? *Journal of Personality and Social Psychology*, *71*, 665–679.

Block, J. H. (1965). The child-rearing practices report. Berkeley Institute of Human Development, University of California, Berkeley.

Bradburn, N. M. (1983). Response effects. In P. Rossi, J. Wright, & A. Anderson (Eds.), *Handbook of survey research* (pp. 289–328). New York: Academic Press.

Brewer, W. F. (1988). Memory for randomly sampled autobiographical events. In U. Neisser & E. Winograd (Eds.), *Remembering considered: Ecological and traditional approaches to the study of memory* (pp. 21–90). Cambridge, UK: Cambridge University Press.

Burton, S., & Blair, E. (1991). Task conditions, response formulation processes, and response accuracy for behavioral frequency questions in surveys. *Public Opinion Quarterly*, *55*, 50–79.

Cannell, C., Miller, P., & Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology 1981* (pp. 389–437). San Francisco: Jossey-Bass.

Catania, J. A., Binson, D., Canchola, J., Pollack, L. M., & Hauck, W. (1996). Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly*, 60, 345–375.

Chu, A., Eisenhower, D., Hay, M., Morganstein, D., Neter, J., & Waksberg, J. (1992). Measuring the recall error in self-reported fishing and hunting activities. *Journal of Official Statistics*, 8, 19–39.

Clayman, M. L., & Wissow, L. S. (2004). Pediatric residents' response to ambiguous words about child discipline and behavior. *Patient Education and Counseling*, 55, 16–21.

Conrad, F. G., & Blair, J. (1996). *From impressions to data: Increasing the objectivity of cognitive interview*. Paper presented at the 1996 Joint Statistical Meetings, Chicago, August 4, 1996. [As cited in Tourangeau, Rips, and Rasinski, 2000.]

Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly*, 64, 1–28.

Dekovic, M., Janssens, J. M., & Gerris, J. R. (1991). Factor structure and construct validity of the Block Child Rearing Practices Report. *Psychological Assessment*, 3, 182–187.

Del Boca, F. K., & Doll, J. A. (2000). Truth or consequences: The validity of self-report data in health services research on addiction. *Addiction*, 95, 347–360.

DeMaio, T. J., & Rothgeb, J. M. (1996). Cognitive interviewing techniques in the lab and in the field. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.

DiFranceisco, W., McAuliffe, T. L., & Sikkema, K. J. (1998). Influences of survey instrument format on the reliability of self-reported high risk sexual behavior. *AIDS and Behavior*, 2, 329–337.

Fagot, B. L. (1992). Assessment of coercive parent discipline. *Behavioral Assessment*, 14, 387–406.

Fendrick, M., & Kim, J. (2001). Multiwave analysis of retest artifact in the National Longitudinal Study of Youth drug use. *Drug and Alcohol Dependence*, 62, 239–253.

Fowler, F. (1992). How unclear terms affect survey data. *Public Opinion Quarterly*, 56, 218–231.

Fowler, F. J., & Cannel, C. F. (1996). Using behavioral coding to identify cognitive problems with survey questions. In N. Schwarz & S. Sudman (Eds.), *Answering Questions: Methodology for determining cognitive and communicative processes in survey research* (pp. 15–36). San Francisco: Jossey-Bass.

Goldstein, K. P., Kviz, F. J., & Daum, R. S. (1993). Accuracy of immunization histories provided by adults accompanying preschool children to a pediatric emergency department. *JAMA*, 270, 2190–2194.

Groves, R. M., & Kahn, R. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.

Handlin, A. H. (1994). Responses and order effects in referendum voting: Exploring the influence of contextual bias on public policy. *Journal of Business Research*, 30, 95–109.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 43, 447–462.

Holden, G. W. (2001). Parenthood. In J. Touliatos, B. F. Perlmutter, & M. A. Straus (Eds.), *Handbook of family measurement techniques* (Vol. 1, pp. 137–149). Thousand Oaks, CA: Sage.

Jabine, T., Straf, M., Tanur, J., & Tourangeau, R. (1984). *Cognitive aspects of survey methodology: Building a bridge between disciplines*. Washington, DC: National Academy Press.

Jacob, T., Moser, R. P., Windle, M., Loeber, R., & Stouthamer-Loeber, M. (2000). A new measure of parenting practices involving preadolescent- and adolescent-aged children. *Behavior Modification*, 24, 611–634.

Jenkins, C., & Dillman, D. (1997). Towards a theory of self-administered questionnaire design. In L. Lyberg, P. Biermer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 165–196). New York: Wiley.

Jo, M. S. (2000). Controlling social-desirability bias via method factors of direct and indirect questioning in structural equation models. *Psychology and Marketing*, 17, 137–148.

Johnson, T., Houghland, J., & Clayton, R. (1989). Obtaining reports of sensitive behaviors: A comparison of substance use reports from telephone and face-to-face interviews. *Social Science Quarterly*, 70, 174–183.

Kovar, M. G., & Royston, P. (1990). Comment on Suchman and Jordan. *Journal of the American Statistical Association*, 85, 246–247.

Krevans, J., & Gibbs, J. C. (1996). Parents' use of inductive discipline: Relations to children's empathy and prosocial behavior. *Child Development*, 67, 3263–3277.

Krosnick, J. A., & Alwin, D. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51, 201–219.

Lessler, J. T., Tourangeau, P., & Salter, W. (1989). *Questionnaire design in the cognitive research laboratory: Results of an experimental prototype.* Vital and Health Statistics, Series 6, No. 1 (DHHS Pub. No. PHS 89-1076). Washington, DC: US Government Printing Office. [As cited in Tourangeau, Rips, and Rasinski, 2000.]

Locke, L. M., & Prinz, R. J. (2002). Measurement of parental discipline and nurturance. *Clinical Psychology Review*, 22, 895–929.

Loftus, E. F. (1996). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.

Lovejoy, M. C., Weis, R., O'Hare, E., & Rubin, E. C. (1999). Development and initial validation of the Parent Behavior Inventory. *Psychological Assessment*, 11, 534–545.

Macoby, E. E., & Martin, J. (1983). Socialization in the context of the family: Parent–child interaction. In P. H. Mussen (Series Ed.) & E. M. Hetheringon (Vol. Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development* (pp. 1–101). New York: Wiley.

Martin, J., O'Muircheartaigh, C., & Curtice, J. (1993). The use of CAPI for attitude surveys: An experimental comparison with traditional methods. *Journal of Official Statistics*, 9, 641–661.

Means, B., & Loftus, E. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology*, 5, 297–318.

Means, B., Nigam, A., Zarrow, M., Loftus, E., & Donaldson, M. (1989). Autobiographical memory for health-related events. *Vital and Health Statistics*, Series 6, No. 2 (DHHS Pub No. (PHS) 89-1077). Washington, DC: US Government Printing Office.

Means, B., Swan, G. E., Jobe, J. B., & Esposito, J. L. (1994). The effects of estimation strategies on the accuracy of respondents' reports of cigarette smoking. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports*. New York: Springer-Verlag.

Menon, G. (1996). Are the parts better than the whole? The effects of decompositional questions on judgments of frequent behavior. *Journal of Marketing Research*, 34, 335–346.

Menon, G., & Yorkston, E. A. (2000). The use of memory and contextual cues in the formation of behavioral frequency judgments. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 3–7). Mahwah, NJ: Earlbaum.

Milner, J. (1994). Assessing physical abuse risk: The child abuse potential inventory. *Child Psychology Review*, 14, 547–583.

Morsbach, S. K., & Prinz, R. J. (2004, May). *An experimental study of bias in parental reports about parenting*. Poster presented

at the annual meeting of the Society for Prevention Research, Quebec City, Canada.

Mosher, W. D., & Duffer, Jr., A. P. (1994). *Experiments in survey data collection: The National Survey of Family Growth pretest.* Paper presented at the meeting of the Population Association of American, May 1994, Miami, FL. [As cited in Tourangeau, Rips, & Rasinski, 2000.]

Moxey, L. M., & Sanford, A. J. (1993). *Communicating quantities.* Hillsdale, NJ: Earlbaum.

Neter, J., & Waksberg, J. (1964). A study of response errors in expenditures from household interviews. *Journal of the American Statistical Association*, *59,* 18–55.

Ollendick, T. H., & Hersen, M. (1984). *Child behavioral assessment: Principles and procedures.* New York: Pergamon Press.

Ong, A. D., & Weiss, D. J. (2000). The impact of anonymity on responses to sensitive questions. *Journal of Applied Social Psychology*, *30,* 1691–1708.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing, and attitude measurement.* London: Pinter.

Paulhaus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.

Perepletchikova, F., & Kazdin, A. (2004). Assessment of parenting practices related to conduct problems: Development and validation of the Management of Children's Behavior Scale. *Journal of Child and Family Studies*, *13,* 385–403.

Peterson, L., Tremblay, G., Ewigman, B., & Popkey, C. (2002). The Parental Daily Diary: A sensitive measure of the process of change in a child maltreatment prevention program. *Behavior Modification*, *26,* 627–647.

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? In P. V. Marsden (Ed.), *Sociological methodology* (Vol. 24, pp. 73–104). Beverly Hills, CA: Sage.

Presser, S., & Stinson, L. (1998). Data collection mode and social desirability bias in self-reported religious attendance. *American Sociological Review*, *63,* 137–145.

Rasinski, K. A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly*, *53,* 388–394.

Reid, J. B., Patterson, G. R., & Snyder, J. (Eds.). (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention.* Washington, DC: American Psychological Association.

Reitman, D., Rhode, P. C., Hupp, S. D. A., & Altobello, C. (2002). Development and validation of the parental authority questionnaire—Revised. *Journal of Psychopathology and Behavioral Assessment*, *24,* 119–127.

Schaeffer, N. C. (1991). Hardly ever or constantly? Group comparisons using vague quantifiers. *Public Opinion Quarterly*, *55,* 395–423.

Schaeffer, N. C. (2000). Asking questions about threatening topics: A selective overview. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 105–121). Mahwah, NJ: Earlbaum.

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, *61,* 576–602.

Schober, M. F., Conrad, F. G., & Fricker, S. S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology*, *18,* 169–188.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question form, word, and context.* New York: Academic Press.

Schwarz, N. (1999a). Self-reports: How the questions shape the answers. *American Psychologist*, *54,* 93–105.

Schwarz, N. (1999b). Self-reports of behaviors and opinions: Cognitive and communicative processes. In N. Schwarz & D. Park (Eds.), *Cognition, aging, and self-reports* (pp. 17–43). Hove, UK: Psychology Press/Erlbaum.

Schwarz, N., Hippler, H., & Noelle-Neumann, E. (1991). A cognitive model of response-order effects in survey measurement. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 187–201). New York: Springer-Verlag.

Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response categories: Effects on behavioral reports and comparative judgments. *Public Opinion Quarterly*, *49,* 388–395.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22,* 127–160.

Sessa, F. M., Avenevoli, S., Steinberg, L., & Morris, A. S. (2001). Correspondence among informants on parenting: Preschool children, mothers, and observers. *Journal of Family Psychology*, *15,* 53–68.

Shelton, K. K., Frick, P. J., & Wootton, J. (1996). Assessment of parenting practices in families of elementary school-age children. *Journal of Clinical Child Psychology*, *25,* 317–329.

Simons, R. L., Whitbeck, L. B., Conger, R. D., & Wu, C. (1991). Intergenerational transmission of harsh parenting. *Developmental Psychology*, *27,* 159–171.

Singer, E., Von Thurn, D. R., & Miller, E. R. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *The Public Opinion Quarterly*, *59,* 66–77.

Smith, T. W. (1992). Discrepancies between men and women in reporting number of sexual partners: A summary from four countries. *Social Biology*, *39,* 203–211.

Stoff, D. M., Breiling, J., & Maser, J. D. (Eds.). (1997). *Handbook of antisocial behavior.* New York: Wiley.

Suchman, L., & Jordan, B. (1992). Validity and the collaborative construction of meaning in face-to-face surveys. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys.* New York: Sage.

Sudman, S., Finn, A., & Lannom, L. (1984). The use of bounded recall procedures in single interview. *Public Opinion Quarterly*, *48,* 520–524.

Tanur, J. M. (Ed.). (1991). *Questions about questions: Inquiries into the cognitive bases of surveys.* New York: Sage.

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 3–7). Mahwah, NJ: Earlbaum.

Tourangeau, R. (2004, June). *The psychology of survey response.* Lecture presented at the Summer Institute in Survey Research Techniques, Ann Arbor, MI.

Tourangeau, R., Rasinski, K., Jobe, J. B., Smith, T. W., & Pratt, W. (1997). Sources of error in a survey of sexual behavior. *Journal of Official Statistics*, *13,* 341–365.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response.* Cambridge, UK: Cambridge University Press.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60,* 275–304.

Turner, C. F., Lessler, J. T., & Devore, J. (1992). Effects of mode of administration and wording on reporting of drug use. In C. Turner, J. Lessler, & J. Gfoerer (Eds.), *Survey measurement of drug use: Methodological studies* (pp. 177–220). Rockville, MD: National Institute on Drug Abuse.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology*, *18,* 225–252.

Willis, G. (1997). *NCHS Cognitive Interviewing Project: General coding scheme for questionnaire problems*. Hyattsville, MD: National Center for Health Statistics. [As cited in Tourangeau, Rips, & Rasinski, 2000.]

Willis, G., Brittingham, A., Lee, L., Tourangeau, R., & Ching, P. (1999). *Response errors in surveys of children's immunizations*. Vital and Health Statistics, Series 6, Number 8. Hyattsville, MD: National Center for Health Statistics.

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Beverly Hills, CA: Sage.

Willis, G. B., & Schechter, S. (1997). Evaluation of cognitive interviewing techniques: Do the results generalize to the field? *Bulletin de Methodologie Sociologique, 55,* 40–66.

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior. *Public Opinion Quarterly, 55,* 613–639.