An Exploratory Study about Inaccuracy and Invalidity in Adolescent Self-Report Surveys

Xitao Fan, Brent C. Miller, Kyung-Eun Park, Bryan W. Winward, Mathew Christensen, Harold D. Grotevant and Robert H. Tai

Field Methods 2006 18: 223
DOI: 10.1177/152822X06289161

An Exploratory Study about Inaccuracy and Invalidity in Adolescent Self-Report Surveys

XITAO FAN
University of Virginia
BRENT C. MILLER
Utah State University
KYUNG-EUN PARK
Gyeonggido Family and Women Development Institute
BRYAN W. WINWARD
Utah State University
MATHEW CHRISTENSEN
Colorado Department of Public Health and Environment
HAROLD D. GROTEVANT
University of Minnesota
ROBERT H. TAI
University of Virginia

Using Add Health data, the authors provide evidence that some adolescents gave inaccurate and/or invalid responses on a self-administered questionnaire. Further analyses show that these adolescents were much more likely to report extreme levels on psychosocial and behavioral outcome variables. A distinction was made between inaccurate responders (e.g., inaccurate/false responses due to carelessness or confusion) and jokesters (e.g., intentional false responses). The findings show that the jokesters showed considerably more pronounced distorting effects on some psychosocial and behavioral outcome variables than the inaccurate responders did. The authors suggest that although this jokester effect may not seriously bias the results in studies that focus on large groups, for research focusing on some special subgroups (e.g., adoption groups, immigrant groups, disability groups), this effect could pose a serious challenge for the validity of research findings.

Keywords: *adolescents; self-administered questionnaires; surveys; self-reports; response validity*

Self-report data collection methods (questionnaires and interviews) are widely used in social and behavioral science research. Self-administered questionnaires (SAQs) have been especially popular because compared with

Field Methods, Vol. 18, No. 3, August 2006 223–244
DOI: 10.1177/152822X06289161

telephone or face-to-face interviews, SAQs are inexpensive and do not have interviewer bias. Furthermore, an SAQ is more likely than an interview to provide anonymity and privacy, encouraging more candid responses on sensitive issues (Babbie 1995:chap. 10), and less likely to induce socially desirable responses (Fujii, Hennessy, and Mak 1985). On the other hand, interviews generally result in more complete data collection, fewer misunderstood questions, and higher response rates (Babbie 1995:chap. 10).

Survey methods in social science research, however, have long been associated with concerns about data quality and response validity (for a collection of issues, see Stone et al. 2000). Because response validity is a concern for social scientists in general, researchers in different disciplines have long studied the issues related to response validity in self-reports. More than two decades ago, Bernard et al. (1984) provided a comprehensive literature review about the problem of informant accuracy. Their review of studies in a variety of research areas (e.g., child care, health care, communication and social interactions, and many other studies in different areas) indicated that informant inaccuracy was a common problem that cut across a wide spectrum of social science disciplines. Based on their review, Bernard et al. concluded that "on average, about half of what informants report is probably incorrect in some way" (p. 503). They further lamented that despite the evidence, this issue did not seem to have penetrated graduate training or social science research and that "informant inaccuracy remains both a fugitive problem and a well-kept open secret" (p. 504).

In situations in which survey respondents are adolescents and in which an SAQ is completed in a classroom or peer group setting, researchers may be especially concerned about the validity of survey responses. In a classroom situation with other students as peers, there may be an elevated chance that some respondents will answer survey questions inaccurately. Spirrison, Gordy, and Henley (1996) compared the response validity between in-class and after-class respondents and showed a variety of response differences between the two groups on SAQ questions. In their study, the SAQ contained

We gratefully acknowledge the support provided by grant HD36479 from the National Institute of Child Health and Human Development, to Brent C. Miller, principal investigator. Inquiries about this article can be sent to Xitao Fan (xfan@virginia.edu). This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris and funded by grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from seventeen other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for their assistance in the original design. Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (www.cpc.unc.edu/addhealth/contract.html).

a validity subscale to help the researchers identify inconsistent/invalid responses. Their findings showed that the built-in validity scale flagged significantly more invalid in-class respondents than after-class respondents, suggesting that unsupervised in-class SAQs in school settings may be more vulnerable to invalid or less valid responding.

Researchers have used various approaches to detect inaccuracies and reduce the problem of response inaccuracy/invalidity in self-reports in general and adolescent self-reports in particular. There is a long and voluminous research history about these issues. Sudman and Bradburn (1974, 1982) reviewed the various kinds of threats to response validity and offered some practical suggestions and guidelines for avoiding or reducing these threats. Many of these suggestions are still being echoed today (e.g., Babbie 1995). For example, researchers collecting longitudinal data can ask the same questions on multiple occasions, thus providing an opportunity to identify inconsistent or incorrect responses. This method can be helpful when a behavior, such as sexual activity, is unlikely to be externally validated. In some cases, researchers can compare SAQ results with national parameters to gauge response consistency. Lauritsen and Swicegood (1997) used these methods to study early sexual experiences of adolescents, and they estimated that over a 7-year period, 32% of adolescent respondents in their study were inconsistent in reporting when their first sexual experience occurred.

Another approach for detecting SAQ inaccuracy uses valid and reliable external measures (physical measurements) against which SAQ responses can be compared (Newcomer and Udry 1988). Researchers have administered SAQs to teenagers and then given them physical examinations to determine how well SAQ responses matched physical measures. In one such study, Goodman, Hinden, and Khandelwal (2000) found that 47% of teens reported themselves to be very overweight when they were not. However, teens may be overly sensitive to the overweight issue because of cultural/societal expectations. As a result, such discrepancies between self-reports and physical measurements may not be the result of SAQ inaccuracies but rather represent the differences between the teens' psychological expectations about ideal body weight and the physical criteria established for overweight classification. As another example, in a study concerning social desirability in an interview situation, Fujii, Hennessy, and Mak (1985) compared interview responses on energy consumption with actual energy consumption and concluded that the findings supported their hypothesis that respondents tended to provide socially desirable answers (i.e., lower-level energy consumption) during an interview.

Similarly, adolescent substance use (e.g., drugs, alcohol, and smoking) can be determined using biochemical measures in blood, urine, or saliva.

Such measures taken at the same time that SAQs are administered can serve as external criteria for detecting SAQ response invalidity (see, e.g., Akers et al. 1983; Hser 1993; Williams and Nowatzki 2005). Research in these areas suggests that about 20% to 30% of high school seniors falsely state that they have not used drugs. These studies also found that as society becomes less tolerant of drug abuse, there is an increased tendency for teenagers to underreport drug use in their SAQ responses (Hser 1993). Messina et al. (2000) reported similar findings for adults in therapeutic settings.

In recent years, many studies have investigated the effects of using technology in interview/survey administration. In general, the purpose of using technology is to provide more privacy in the self-reporting process, with the expectation that this will enhance self-reports about sensitive information. Turner et al. (1998) used SAQs and followed up with either personal interview or audio computer-assisted self-interviewing method (CASI) to compare responses of individuals to similar questions. They concluded that the more privacy the interviewing mode provides, the more likely that respondents will provide answers to sensitive questions related to topics such as sexual behavior, same-gender attraction, abortion, and drug use and the less likely that they will give responses such as "I don't know" or "refuse to answer."

Liu and Detels (1999) investigated the response validity in self-reporting of sexual behaviors using similar technology. By comparing the same respondents' answers to related questions and comparing sexual activities reported by husband and wife within a couple, they concluded that privacy technologies (tape recorder, earphones, etc.) enhanced the validity of responses to sensitive survey questions. Similarly, Paschall, Ornstein, and Flewelling (2001) concluded that audiocassette or audio CASI interviewing methods could yield self-report measures about involvement in the juvenile justice system with a fairly high degree of criterion validity. Supple, Aquilino, and Wright (1999), Tourangeau and Smith (1996), and many others reported similar findings.

Some researchers have investigated the issue of adolescent response validity by examining the discrepancy between adolescents' self-reports and the responses from others, such as parents, siblings, or friends. This approach can help identify inconsistencies and verify information concerning private and sensitive issues such as adoption, race, socioeconomic status, and drug and alcohol use. These studies suggest that between 18% and 28% of adolescents give invalid SAQ responses, depending on the sensitivity of the questions asked (Hser 1993; Ensminger et al. 2000; Miller et al. 2001).

Efforts have been made to identify respondents who may have provided grossly incorrect responses. A very popular approach is to embed a "validity

scale" (e.g., "fake scale" or "lie scale") in a self-report questionnaire to detect if a respondent is either "faking good" (also known as "social desirability response set") or "faking bad" (Anastasi and Urbina 1997; Worthen et al. 1999). This approach of embedding a validity scale in a self-report questionnaire has been used in some standardized measures, for example, the Minnesota Multiphasic Personality Inventory.

Response sets such as faking good should be considered as being different from more intentional mischievous/dishonest responses. For example, it is one thing to try (consciously or subconsciously) to create a favorable impression (social desirability), but it is quite another thing for a respondent to state that he or she is handicapped when, in fact, he or she is not or to misrepresent oneself as being an adoptee when the respondent knows that such is not the case. Issues related to mischievous or dishonest response patterns are what we are interested in for this article.

In large-scale surveys, there has been little effort to incorporate design features in survey instruments to help "flag out" responders who may have intentionally provided grossly incorrect responses. In this article, we describe some exploratory efforts to identify adolescent respondents who may have provided dishonest or mischievous responses in a large-scale self-report survey, and we examine the distorting effects of these responders on some substantively meaningful outcome variables.

METHODS

Data Source

We used the National Longitudinal Study of Adolescent Health (Add Health) in our analyses. Add Health is a large and nationally representative data collection project. The data of the first wave of Add Health were collected during 1994 to 1995 to measure social and familial contextual variables that influence the health, well-being, and health-related behaviors of adolescents in grades 7 through 12. A cluster sampling design was used for data collection. Add Health includes data relevant to adolescent health and well-being from school administrators, parents, and adolescents themselves. Data were first collected in schools and later in homes from a subsample.

From September 1994 to April 1995, paper-and-pencil SAQs were administered to 90,118 adolescents (grades 7–12) in school settings. The SAQ questions assessed demographic characteristics, family structure, friendship networks, and not-so-sensitive risk behaviors. Of the 90,118 participants in

the school SAQ sample, a subsample (slightly more than 20,000) was included in the Wave I in-home interview conducted about 1 year after the school SAQ. Many questions were asked about demographics, parents, friends, and risk behaviors (including sensitive and illegal behaviors). A parent (or custodian adult) of each adolescent in the Wave I in-home interview sample was asked to complete an interviewer-assisted questionnaire at home at the same time as Wave I in-home interviews of the adolescent. Approximately 85% of the parents responded to the parent survey. More details about the Add Health study design are available from the National Longitudinal Study of Adolescent Health (http://www.cpc.unc.edu/projects/addhealth). For the analyses in this study, three linked data sets were used: the in-school adolescent SAQ data, Wave I in-home interviews of adolescents (in-home interview), and the parent survey completed at home.

Because of its multiple data sources, Add Health offered a unique opportunity to examine adolescents' response accuracy in SAQ. The multiple data sources provided useful mechanisms for verifying adolescent responses on the Add Health SAQ administered in school settings. This made it possible to detect inaccurate responders on the SAQ, even those who might have mischievously provided incorrect responses intentionally on the Add Health SAQ ("jokesters," so to speak). In the following sections, we describe the process of identifying several groups of inaccurate responders by verification across the Add Health data sets and identify a group of jokesters who gave more systematic inaccurate responses. We also assess the distorting effects of these groups on some substantively meaningful psychosocial outcome variables.

RESULTS

Our previous research efforts related to adoption adjustment issues (see Miller et al. 2001; Fan et al. 2002) alerted us that a small group of adolescents in the in-school SAQ appeared to have (either intentionally or because of confusion) misrepresented themselves as being adoptees when they were not. To explore the extent to which an invalid response pattern existed and could be replicated on other personal variables in the in-school SAQ, we identified several SAQ personal variables, the validity of which could be verified through related Add Health data sets (i.e., Wave I in-home interviews and Wave I parent surveys). There were two criteria for selecting these variables: (1) These should be personal variables that were reasonably easy to understand by adolescents, and (2) the same (or very similar) variables existed in the other two related Add Health data sets such that the

response validity on the SAQ could be triangulated. We were also interested in understanding (1) the extent to which adolescent misrepresentation occurred on multiple personal variables and (2) if such personal misrepresentation was related to invalid responses on other substantively meaningful psychosocial-behavioral variables.

DESCRIPTION OF INACCURATE RESPONDERS

Defined by Adoption Status

As described in detail in Miller et al. (2001) and Fan et al. (2002), in the school SAQ, the self-claimed adolescent adoptees could be classified based on two questions: (1) "Are you adopted?" and (2) "Do you live with either of your biological parents?" Adolescents who answered "yes" to the first and "no" to the second question were classified as adoptees. In the in-home interview, the adoption status was derived from the home living arrangement reported to the interviewer by the adolescent. As described in Miller et al. (2001), from all the people living in the household of the respondent, if respondents stated they were living with both adoptive father and adoptive mother, or only adoptive father without mother, or only adoptive mother without father, the respondent was classified as an adoptee. In the parent survey, the adoption status of an adolescent was determined if the parent responded that (1) he or she was the adoptive father/mother and (2) no biological parent(s) of the target child lived in the household.

Miller et al. (2001) gave detailed accounts for the discrepancies across the three Add Health data sets in defining the adoption status of the adolescents. Among other things, based on the triangulation of the three data sets, three groups of adolescents were identified with regard to adoption status: true adoptees (n = 370), true nonadoptees (n = 14,662), and false adoptees (n = 88). It is reasonable to conclude that the true adoptees and true nonadoptees provided honest responses to the in-school SAQ because their responses were consistent with their home living arrangements in the in-home interview and were confirmed by their parents during the parent survey. On the other hand, the false adoptees in the school SAQ were contradicted by both their home living arrangements in the in-home interview and by their parents' responses in the parent survey. This group was considered as inaccurate responders. In the school SAQ, thirteen adolescents reported themselves as nonadoptees, but their later home living arrangements and their parent report indicated that they were really adoptees. However, we are less certain to call these adolescents (n = 13) false nonadoptees because

these young adoptees might not have been informed that they were adoptees when they were reporting in the school SAQ. For this reason, we did not consider this group as inaccurate responders.

Defined by "Born in United States" Question

A second group of responders was defined based on the inconsistency pattern on the questions in the three Add Health data sets about whether they were "born in the United States." A total of 11,550 cases were consistently identified as born in the United States, and 863 were consistently identified as being "not born in the United States" in all three data sources. On the other hand, 176 individuals answered that they were not born in the United States in the SAQ but later reported that they were born in the United States in the face-to-face in-home interview, and their parent/guardian's response confirmed that they were born in the United States. These inconsistency patterns led us to believe that some adolescents were either not serious in answering this question on the SAQ or incorrectly marked that they were not born in the United States on the SAQ. We labeled the three groups described above as "true U.S. born" (n = 11,550), "true non-U.S. born" (n = 863), and "false non-U.S. born" (n = 176). Twenty-six adolescents self-reported to be U.S. born in the school SAQ, but their later responses in the home interview and their parents' report indicated that they were not born in the United States. Again, we consider it likely that that some adolescents might not have been informed about where they were born; as a result, they might have assumed that they were born in the United States. For this reason, we did not consider this group (n = 26) to be inaccurate responders.

Defined by "Using Artificial Limb" Question

The third group of inaccurate responders was composed of those answering questions in the SAQ and the in-home interview about whether they had an artificial limb. Of the 15,356 SAQ sample members included in the Wave I in-home interview sample, 253 respondents stated that they had used an artificial limb (hand, arm, leg, or foot) for the past year or more, indicating a permanent physical disability. However, when interviewed later in the Wave I in-home interview, only 2 of these 253 adolescents reported that they were using an artificial limb; the overwhelming majority (248) reported not using any artificial limb, and 3 did not answer this question.

The artificial limb question was not included in the Wave I parent survey. However, because the use of an artificial limb suggests a permanent physical disability and could be verified quite easily during the face-to-face

interview, it is unlikely that the adolescents would have provided erroneous information during the face-to-face in-home interview. For this reason, in the absence of parent response, the in-home interview data offer a sufficient validity check for the information provided in the SAQ. This finding suggests that most adolescents previously reporting using an artificial limb on school SAQs were probably not truthful about this question. For us, this represents another group of inaccurate responders who might not have taken the SAQ survey seriously. In the school SAQ, ten adolescents reported not using an artificial limb but later reported using an artificial limb in the home interview. This group was not considered inaccurate responders because it is possible that some physical disability condition could have occurred for some adolescents between the school SAQ and the home interview

SUPPORTING EVIDENCE FOR IDENTIFICATION OF INACCURATE RESPONDERS

The three groups of inaccurate responders defined above also showed a higher degree of inconsistency on some other salient questions such as gender status, age, and ethnicity/race. This section describes the inconsistency pattern on these variables as supporting evidence for the classification of these three groups of inaccurate responders.

Gender Status Inconsistency

The consistency between self-reported gender status in the SAQ and the in-home interviewer's gender classification by the interviewer was evaluated. The inaccurate responder groups identified above showed considerably higher inconsistency rates. The true responders (i.e., those who were not classified into any of the three inaccurate responder groups) have almost perfect consistency rates (>99%) in gender classification. However, among the false adoptee group, 37% of the self-identified females in the SAQ were later classified as males by the interviewers. For the false non-U.S. born group, 31% of the self-reported females in the SAQ were later classified as males by the interviewers. Among the false artificial limb group, 34% of the self-reported females were later classified as males by the interviewers. The overwhelming majority of the inaccurate responder groups were classified as males by the interviewers (80%, 77%, and 84%, respectively, for the false adoptees, the false non-U.S. born, and the false artificial limb group), and it appeared that many of these male respondents "faked" their gender, claiming to be females in the SAQ.

Age Inconsistency

The school SAQ contains self-reported age. In the Wave I in-home interview, the respondent's age can be computed by subtracting the self-reported birth year from the year of interview. It is expected that the correlation between the self-reported age in the SAQ and the computed age in the Wave I in-home interview should be very high. This expectation was confirmed for most SAQ participants but not for the three groups of inaccurate responders. The correlation coefficients between the SAQ and in-home interview ages typically ranged from .95 to .97 for true responders but were substantially lower for the inaccurate responder groups, with correlation coefficients being .47, .41, and .47, respectively, for the false adoptees, the false non-U.S. born group, and the false artificial limb group.

Ethnicity/Race Inconsistency

In both the SAQ and the in-home interview, the adolescents reported their ethnicity/race to be one of six categories: White, Black, Asian/Pacific Islander, American Indian, Hispanic, and Other. Consistency was checked between the two data sets. For the true responders, the consistency rates were about 80%. But for the three groups of inaccurate responders, the consistency rates were 47%, 36%, and 47%, respectively. Ethnicity/race is more complicated than gender and age and could be confusing for many adolescents, so lower consistency across data sets is expected. But the much lower consistency for the inaccurate responder groups suggests that some of them might not have responded to the questions truthfully.

All the consistency patterns described above suggest that the groups of inaccurate responders were less consistent in their responses to other questions (i.e., gender, age, and ethnicity/race). These patterns provide supporting evidence that inaccurate responders may be different from true responders in that they were either inaccurate/careless/confused in responding to the questions or that they provided mischievous/false answers to various questions.

DESCRIPTION OF JOKESTERS

Although we identified several groups of inaccurate responders, the reasons for their inaccurate responses may be different: They could be due to carelessness, confusion, lack of efforts, or intentional mischief. Although an adolescent was classified as an inaccurate responder, it would be premature to conclude that he or she provided inaccurate responses intentionally. On the

other hand, if a respondent was classified into multiple groups of inaccurate responders, the evidence became stronger that the respondent might have intentionally or mischievously provided erroneous responses. In this case, it is probably appropriate to call these respondents jokesters, indicating that the degree of their inaccuracy on the SAQ was probably more than what we would expect from carelessness or confusion. We used this rationale to further narrow down a smaller group of jokesters among the inaccurate responders described previously.

Four hundred eighteen adolescents were classified into at least one of the three groups of inaccurate responders (based on questions on adoption, whether they were born in the United States, and whether they had an artificial limb). Of these, 336 belonged to only one group of inaccurate responders (either false adoptee, false non-U.S. born, or false artificial limb group). We gave these adolescents the benefit of the doubt and considered them "inaccurate responders," meaning that we are not certain about the reason for their inaccurate response on the question of interest. Another seventy were flagged as belonging to two of the three groups of inaccurate responders, and twelve were classified as belonging to all three groups of inaccurate responders. Because of sample size consideration, we collapsed the latter two into one group ($n = 82$). As these adolescents belong to multiple groups of inaccurate responders, the evidence was stronger that they might have provided inaccurate responses intentionally on the SAQ. We call this group jokesters.

POTENTIAL IMPACT OF INACCURATE RESPONSES ON SUBSTANTIVE RESEARCH

Now that we have identified several groups of inaccurate responders and one group of jokesters in the school SAQ data based on the triangulation from multiple data sources, a question that logically follows is, Do these inaccurate responders and jokesters introduce enough systematic data error to have a significant negative impact on substantive research? The following sections describe how these groups can potentially distort substantive findings.

Effect of Inaccurate Responders on Adoption

To what extent could the false adoptees affect the comparison between adoptee group and nonadoptee group on substantively meaningful outcome variables? Table 1 presents two comparisons on a group of SAQ outcome variables. These outcome variables are the typical school, psychosocial, and

TABLE 1
Effect of Inaccurate Responders Defined by Adoption Status ("False Adoptees")

Outcome Variable	*True Adoptees* (n = *376) vs. Nonadoptees* (n = *14,662)*	*False Adoptees* (n = *88) vs. Nonadoptees* (n = *14,662)*
School grades (+)	–0.11	–0.24
School troubles	–0.10	0.53
Positive school feelings (+)	0.05	–0.99
Skipping school	0.02	1.97
Smoking	0.05	0.94
Drinking	–0.03	1.45
Drunk	–0.02	1.68
Self-esteem (+)	–0.01	–0.96
Emotional distress	0.12	1.02
Future hope (+)	–0.01	–1.36
Health problems	0.06	1.08
Physical problems	0.06	1.73
Sickness	–0.08	1.15
Fighting	0.05	1.36
Lie to parents	–0.11	0.84
Mean of absolute effect size	0.06	1.09

NOTE: + = A positive variable on which a higher value is desired. The column entries are effect sizes in the form of the standardized mean difference between the two groups $[(\overline{X}_{1st\ Group} - \overline{X}_{2nd\ Group})/(SD_{2nd\ Group})]$. Classification for the groups (true adoptees, jokester adoptees, and nonadoptees) is based on triangulation from three data sources as described in the text. The sample sizes for the adopted and nonadopted groups presented in the table are the maximum sample sizes for the groups in this analysis. The actual sample sizes for a particular outcome variable comparison may be lower because of missing data. Because only selected groups were used, the sampling weight was not applied in these analyses.

behavioral variables commonly used in adolescent research and were not specifically selected for this article. In fact, all these variables have been used in previous research studies (Miller, Fan, Christensen, Coyl, et al. 2000; Miller, Fan, Christensen, Grotevant, et al. 2000; Fan et al. 2002). Detailed discussions about the construction of these variables and their psychometric characteristics are available from the previously published studies.

The first comparison is between the true adoptees and the true nonadoptees. Column 1 shows that there is little difference between the two groups on these outcome variables, as indicated by the very small effect sizes between the two groups' means (mean absolute effect size = 0.06). The second comparison is between the false adoptees and the true nonadoptees. This second column shows that the false adoptee group responded quite differently from the true nonadoptee group, as shown by

TABLE 2
Effect of Inaccurate Responders Defined by "Born in United States" Question ("False Non-U.S. Born")

Outcome Variable	*True Non-U.S. Born* (n = *863*) *vs. True U.S. Born* (n = *11,550*)	*False Non-U.S. Born* (n = *176*) *vs. True U.S. Born* (n = *11,550*)
School grades (+)	–0.03	–0.32
School trouble	0.03	0.30
Positive school feelings (+)	0.01	–0.64
Skipping school	0.02	1.33
Smoking	–0.29	0.63
Drinking	–0.24	1.07
Drunk	–0.26	1.24
Self-esteem (+)	0.01	–0.49
Emotional distress	–0.10	0.60
Future hope (+)	–0.03	–1.32
Health problems	–0.36	0.67
Physical problems	–0.15	1.62
Sickness	–0.04	0.90
Fight	–0.21	1.15
Lie to parents	–0.25	0.59
Mean of absolute effect size	0.13	0.80

NOTE: + = A positive variable on which a higher value is desired. The column entries are effect sizes in the form of the standardized mean difference between the two groups $[(\overline{X}_{\text{1st Group}} - \overline{X}_{\text{2nd Group}})/(SD_{\text{2nd Group}})]$. Classification for the groups is based on triangulation from three data sources as described in the article. The sample sizes for the groups presented in the table are the maximum sample sizes for the groups in this analysis. The actual sample sizes for a particular outcome variable comparison may be lower because of missing data. Because only selected groups were used, the sampling weight was not applied in these analyses.

the generally large effect sizes on many of these outcome variables (mean absolute effect size = 1.09). The results as shown in the second column led to our incorrect interpretation about the comparison between adoptee and nonadoptee groups in previous research (Miller, Fan, Christensen, Grotevant, et al. 2000; Fan et al. 2002).

Effect of Inaccurate Responders on U.S.-Born Question

What distorting effect would the false non-U.S.-born group have on researchers interested in comparisons between U.S.-born and non-U.S.-born groups? In Table 2, the first comparison is between the true non-U.S.-born group and the true U.S.-born group. Column 1 shows little difference

between the two groups on these outcome variables, as indicated by the small effect sizes between the two group means (mean absolute effect size = 0.13). The second comparison is between the false non-U.S.-born group and the true U.S.-born group. Column 2 shows that the false non-U.S.-born group responded quite differently from the true U.S.-born group, as shown by the generally large effect sizes on many of these outcome variables (mean absolute effect size = 0.80). For example, the false non-U.S.-born group was much more likely to report skipping school, drinking, getting drunk, having physical problems, and being involved in fighting.

Effect of Inaccurate Responders on the Artificial Limb Question

Did the inaccurate responders defined by the artificial limb question respond differently than true responders on substantively meaningful outcome variables? Table 3 compares outcome variable means between the true responders on the artificial limb question and the inaccurate responders on this question (n = 248). As is obvious from the table, responses of the two groups on these outcome variables are quite different, as shown by the very large effect sizes on most of the variables (last column), with the mean absolute effect size being 1.21. Again, the inaccurate responders were much more likely to report higher values on negative outcome variables (e.g., skipping school, drinking, and having physical disabilities) but lower values on positive outcome variables (e.g., school grades, self-esteem).

Comparing the Distorting Effects of Inaccurate Responders and Jokesters

Based on the number of inaccurate responder groups into which an adolescent was classified, we distinguished between inaccurate responders (belonging only to one group of inaccurate responders) and jokesters (belonging to multiple groups of inaccurate responders). Is such a distinction meaningful in terms of the distorting effects on some substantively meaningful outcome variables? Table 4 presents the comparison of the estimated distorting effects of the two groups.

Column 1 is the comparison between the inaccurate responders and the true responders, showing that there is some difference between the two groups on these outcome variables, as indicated by some moderate to large effect sizes between the two groups' means on the variables (mean absolute effect size = 0.73). The second column is the comparison between the jokesters and the true responders, showing not only that the responses from the jokesters differed from the true responders but also that the differences were huge or dramatic on many of these outcome variables (mean absolute effect size = 1.89). These two comparisons suggest that the distinction between the inaccurate

TABLE 3
Effect of Inaccurate Responders Defined by "Artificial Limb" Questions on Self-Administered Questionnaires and In-Home Interviews

	Inaccurate Responders on Artificial Limb Questions (n = 248)		*Accurate Responders on Artificial Limb Questions (n = 14,153)*		
Outcome Variable	M	SD	M	SD	*Effect Size*[a]
School grades (+)	2.67	1.05	2.82	0.78	–0.19
School trouble	2.24	1.42	1.59	1.13	0.58
Positive school feeling (+)	2.89	1.32	3.55	0.84	–0.79
Skipping school	3.15	2.70	0.61	1.22	2.08
Smoking	3.15	2.76	1.13	1.99	1.02
Drinking	3.64	2.56	1.18	1.49	1.65
Drunk	1.75	1.32	0.43	0.74	1.78
Self-esteem (+)	3.35	1.31	3.89	0.69	–0.78
Emotional distress	1.84	1.42	1.01	0.79	1.05
Future hope (+)	4.11	2.18	6.31	1.19	–1.85
Health problems	2.23	1.34	1.41	0.68	1.21
Physical problems	1.61	1.58	0.23	0.49	2.82
Sickness	2.35	1.19	1.59	0.68	1.12
Fight	2.39	1.64	0.75	1.09	1.50
Lie to parents	3.73	2.58	2.12	1.78	0.90
Mean of absolute effect size					1.21

NOTE: + = A positive variable on which a higher value is desired. The sample sizes for the groups presented in the table are the maximum sample sizes for the groups in this analysis. The actual sample sizes for a particular outcome variable comparison may be lower because of missing data. Because only selected groups were used, the sampling weight was not applied in these analyses.
a. Effect size = $(\overline{X}_{\text{jokesters}} - \overline{X}_{\text{Nonjokesters}})/SD_{\text{Nonjokesters}}$.

responders and the jokesters had some empirical support, with the jokesters providing some dramatically different responses from the true responders.

SUMMARY AND DISCUSSION

Summary

The findings in this article provide empirical evidence that some adolescents gave inaccurate or invalid responses to questions on a self-administered

TABLE 4
Comparing the Distorting Effects of Inaccurate Responders and Jokesters

Outcome Variable	*Inaccurate Responders (n = 336) vs. True Responders (n = 14,172)*	*Jokesters (n = 82) vs. True Responders (n = 14,172)*
School grades (+)	–0.13	–0.39
School trouble	0.43	0.53
Positive school feeling (+)	–0.49	–1.25
Skipping school	1.17	3.11
Smoking	0.61	1.45
Drinking	0.90	2.51
Drunk	1.01	2.73
Self-esteem (+)	–0.36	–1.28
Emotional distress	0.49	1.68
Future hope (+)	–1.08	–2.58
Health problems	0.59	1.79
Physical problems	1.65	3.54
Sickness	0.66	1.74
Fight	0.90	2.29
Lie to parents	0.42	1.51
Mean of absolute effect size	0.73	1.89

NOTE: + = A positive variable on which a higher value is desired. As explained in the text, inaccurate responders are those classified into only one of the three inaccurate responder groups (adoption, born in the United States, or artificial limb). Jokesters are those classified into two or three inaccurate responder groups. The column entries are effect sizes in the form of the standardized mean difference between the two groups $(\overline{X}_{\text{1st Group}} - \overline{X}_{\text{2nd Group}})/(SD_{\text{2nd Group}})$. Classification for the groups is based on triangulation from three data sources as described in the article. The sample sizes for the groups presented in the table are the maximum sample sizes for the groups in this analysis. The actual sample sizes for a particular outcome variable comparison may be lower because of missing data. Because only selected groups were used, the sampling weight was not applied in these analyses.

questionnaire in school settings. In every comparison between inaccurate responders and true responders, the inaccurate responders reported obviously higher mean scores for problem behaviors and lower mean scores on positive variables. We distinguished between inaccurate responders (inaccurate responses could be due to carelessness or confusion but not necessarily because of intentional mischief) and jokesters (inaccurate responses appeared more likely to be the result of intentional false responses). The findings show that the distorting effects of the jokesters on some common outcome variables in adolescent research were considerably more pronounced than those of the inaccurate responders.

Implications

The analyses and findings presented here suggest that SAQ surveys of adolescents in schools should be viewed cautiously. Large-scale in-school SAQ surveys, such as the Youth Risk Behavior Survey, are likely to have these kinds of invalid data because of the SAQ jokesters' effect shown in these analyses. If the research focus is on large groups, the jokesters' effect should not seriously bias the finding because the number of jokesters is usually very small compared with the total sample. However, if subgroup analyses (e.g., adoption group, immigrant group, disability group) are the research focus, the jokester effect may pose a serious problem for the validity of the findings.

It appears that unusual or exceptional response options tend to attract jokester respondents. In this article, the three questions we analyzed all fit this description. For an average adolescent at school, the answer that "I am adopted," "I was not born in the United States," or "I have an artificial limb" all are rare or unusual responses with some degree of novelty. Although we have limited understanding about what inferences can be made from these observations, we suspect that these unusual or rare response options have a tendency to attract jokesters.

It may appear that this is an implicit argument that interviewing provides better quality data than questionnaires do. This, however, is not the central point of our article. We checked the response validity in the in-school SAQ, and in doing so, we relied on both the interview data and the information from parent surveys. Because the data we used were not collected for the purpose of comparing questionnaire and interview data quality, we do not make general statements about the merits or demerits of the two data collection methods. Instead, the results simply showed that a small number of adolescents may provide invalid responses in an in-school SAQ and that such invalid responses may have nontrivial consequences for analyses involving certain subgroups.

Why do substantial differences exist between the jokesters and the true responders (see Table 4) and, to a lesser degree, between the inaccurate responders and the true responders (see Tables 1, 2, and 3) on the outcome variables we examined? There are two possible explanations. First, the inaccurate responders, and especially the jokesters, might indeed have more problems (psychosocial, behavioral, and academic); this was reflected in their higher scores on the negative outcome variables and lower scores on the positive outcome variables. In other words, the observed substantial differences on the outcome variables are real, and such differences accurately reflect reality, even though the inaccurate responders and the jokesters

provided incorrect/false information on the questions used for defining these groups (i.e., questions about adoption, place of birth, and presence of an artificial limb). Alternatively, it is plausible that the inaccurate responders, and especially the jokesters, not only provided incorrect/false information on the questions used to define these groups but also provided exaggerated and/or false responses to the questions of the outcome variables (e.g., skipping school, drinking, physical problems, and future hopes).

Which of the two hypotheses is closer to the truth? Unfortunately, we are unable to draw a definitive conclusion either way. If some objective indicators for some dependent variables (e.g., school attendance record, actual school grade point average, medical reports on physical/health problems) had been available in the data, it might have been possible to test the two competing hypotheses (jokesters exaggerated problems vs. they indeed had more problems). Without such objective indicators, we will have to live with this uncertainty at this time. But the very large or even dramatic differences on some outcome variables between the jokesters and true responders (see Table 4) led us to believe that the jokesters had probably exaggerated their responses to some extent on some SAQ outcome variables.

Even with the uncertainty about the two competing hypotheses described above, it should be obvious that adolescent jokesters can considerably affect some substantive research findings if this phenomenon is ignored in research practice. For example, the reported differences between adoptees and nonadoptees in the article by Miller, Fan, Christensen, Grotevant, et al. (2000) are very likely to have been inflated by adolescent jokesters on the question of adoption (Fan et al. 2002). Similarly, if researchers interested in issues related to new immigrants made comparisons between U.S.-born and non-U.S.-born adolescents based on an SAQ, the jokester phenomenon would result in inflated differences between the two groups, as shown in Table 2.

Future Directions

It is not possible to prevent a small number of adolescent respondents from providing inaccurate/false information on an SAQ. However, it could be helpful to embed some survey design features in a questionnaire for consistency verification (e.g., a validity scale). Such built-in checks would help SAQ data analysts identify inaccurate responders and jokesters and consider adjustments or deletions in data analyses. Conventional validity checks (e.g., those for flagging social desirability or faking bad) may not be enough for the phenomenon observed in this study. Instead, a more direct approach for checking consistency of factual information may be needed.

Using Add Health in-home interview data, Sieving et al. (2001) developed post hoc validity flags by noting the responses on different items that, when put together, were highly improbable or simply impossible (e.g., a girl who reports in one section having had a child but in another section states that she has never had sex). The feasibility of constructing post hoc validity flags, however, may or may not exist, depending on the type of items included in a questionnaire. It would be much better if this approach is considered in the questionnaire design. For example, related items can be intentionally embedded in different sections of the questionnaire, and the data analyst can construct internal validity checks. Similarly, in longitudinal data collection using SAQs, it is advantageous to include the same or similar measures of nonchangeable characteristics across multiple waves so that validity checks could be constructed to identify jokesters across waves of data collection.

Limitations

Like all research, this study has its own limitations. First, it relied on extant data, which has some inherent constraints. For example, the analysis was limited to the questions available from the data sources, and it is possible that other types of questions might be more likely to solicit invalid responses from adolescents in an SAQ. Second, and more important, after identifying the problem(s) discussed in this article, some general suggestions were offered, but more specific solution(s) to the problem(s) are not yet readily apparent. Some new or innovative strategies may be needed to minimize the problems described in this article and to improve the accuracy of adolescent SAQ responses. Developing such solutions will require more intense experimental studies to test possible strategies or possibilities, and such endeavors are beyond the scope of this article.

REFERENCES

Akers, R. L., J. Massey, W. Clarke, and R. M. Lauer. 1983. Are self-reports of adolescent deviance valid? Biochemical measures, randomized response, and the bogus pipeline in smoking behavior. *Social Forces* 62 (1): 234–51.

Anastasi, A., and S. Urbina. 1997. *Psychological testing*. 7th ed. Upper Saddle River, NJ: Prentice Hall.

Babbie, R. R. 1995. *The practice of social research*. New York: Wadsworth.

Bernard, H. R., P. Killworth, D. Kronenfeld, and L. Sailor. 1984. The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology* 13:495–517.

Ensminger, M. E., C. B. Forrest, A. W. Riley, M. Kang, B. F. Green, B. Starfield, and S. A. Ryan. 2000. The validity of measures of socioeconomic status of adolescents. *Journal of Adolescent Research* 15 (3): 392–410.

Fan, X., B. C. Miller, M. Christensen, B. Bayley, K. E. Park, H. D. Grotevant, M. van Dulmen, and N. Dunbar. 2002. Questionnaire and interview inconsistencies exaggerated differences between adopted and non-adopted adolescents in a national sample. *Adoption Quarterly* 6 (2): 7–27.

Fujii, E. T., M. Hennessy, and J. Mak. 1985. An evaluation of the validity and reliability of survey response data on household electricity conservation. *Evaluation Review* 9 (1): 93–104.

Goodman, E., B. R. Hinden, and S. Khandelwal. 2000. Accuracy of teen and parental reports of obesity and body mass index. *Pediatrics* 106 (1): 52–59.

Hser, Y. 1993. Data sources: Problems and issues. *Journal of Drug Issues* 23 (2): 217–29.

Lauritsen, J. L., and C. G. Swicegood. 1997. The consistency of self-reported initiation of sexual activity. *Family Planning Perspectives* 29 (5): 215–22.

Liu, H., and R. Detels. 1999. An approach to improve validity of responses in a sexual behavior study in a rural area of China. *AIDS & Behavior* 3 (3): 243–49.

Messina, N. P., E. D. Wish, S. Nemes, and B. Wraight. 2000. Correlates of underreporting of post-discharge cocaine use among therapeutic community clients. *Journal of Drug Issues* 30 (1): 119–32.

Miller, B. C., B. K. Bayley, M. Christensen, X. Fan, D. Coyl, H. D. Grotevant, M. van Dulmen, and N. Dunbar. 2001. Who is adopted? Measuring adoption status using national survey data. *Adoption Quarterly* 5 (1): 23–44.

Miller, B. C., X. Fan, M. Christensen, D. Coyl, H. D. Grotevant, and M. van Dulmen. 2000. Adopted adolescents' overrepresentation in mental health counseling: Adoptees' problems or parents' lower threshold for referral? *Journal of the American Academy of Child and Adolescent Psychiatry* 39 (12): 1504–11.

Miller, B. C., X. Fan, M. Christensen, H. D. Grotevant, and M. van Dulmen. 2000. Comparisons of adopted and non-adopted adolescents in a large nationally representative sample. *Child Development* 71 (5): 1458–73.

Newcomer, S., and J. R. Udry. 1988. Adolescents' honesty in a survey of sexual behavior. *Journal of Adolescent Research* 3 (3–4): 419–23.

Paschall, M. J., M. L. Ornstein, and R. L. Flewelling. 2001. African American male adolescents' involvement in the criminal justice system: The criterion validity of self-report measures in a prospective study. *Journal of Research in Crime & Delinquency* 38 (2): 174–87.

Sieving, R. E., T. Beurhing, M. D. Resnick, L. H. Bearinger, M. Shew, M. Ireland, and R. W. Blum. 2001. Development of adolescent self-report measures from the National Longitudinal Study of Adolescent Health. *Journal of Adolescent Health* 28 (1): 73–81.

Spirrison, C. L., C. C. Gordy, and T. B. Henley. 1996. After-class versus in-class data collection: Validity issues. *Journal of Psychology* 130 (6): 635–44.

Stone, A. A., J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtsman, and V. S. Cain, eds. 2000. *The science of self-report: Implications for research and practice*. Mahwah, NJ: Lawrence Erlbaum.

Sudman, S., and N. M. Bradburn. 1974. *Response effects in surveys: A review and synthesis*. Chicago: Aldine.

———. 1982. *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Supple, A. J., W. S. Aquilino, and D. L. Wright. 1999. Collecting sensitive self-report data with laptop computers: Impact on the response tendencies of adolescents in a home interview. *Journal of Research on Adolescence* 9 (4): 467–88.

Tourangeau, R., and T. W. Smith. 1996. Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 60 (2): 275–304.

Turner, C. F., L. Ku, S. M. Rogers, L. D. Lindberg, J. H. Pleck, and F. L. Sonenstein. 1998. Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science* 280 (5365): 867–73.

Williams, R. J., and N. Nowatzki. 2005. Validity of adolescent self-report of substance use. *Substance Use & Misuse* 40 (3): 299–311.

Worthen, B. R., K. R. White, X. Fan, and R. R. Sudweeks. 1999. *Measurement and assessment in schools*. 2nd ed. New York: Allyn & Bacon/Longman.

XITAO FAN is a professor of education in the Curry School of Education, University of Virginia. He is a quantitative methodologist with research interests in quantitative methods, validity and reliability issues in measurement, and self-report response validity. Some recent publications are "Sensitivity of Fit Indices to Misspecified Structural or Measurement Model Components: Rationale of a Two-Index Strategy Revisited" (with A. A. Sivo, Structural Equation Modeling*, 2005) and "Missing Data in Disguise and Implications for Survey Data Analysis" (with L. Wang,* Field Methods*, 2004).*

BRENT C. MILLER is the vice president for research and a professor of family and human development at Utah State University. His research has focused on adolescent sexual behavior and pregnancy and the adjustment of adopted children. Publications include "Teen Childbearing and Public Policy" (with R. Sage and B. W. Winward, in Fragile Families and the Marriage Agenda*, Springer, 2006) and "Family Influences on Adolescent Sexual and Contraceptive Behavior" (*Journal of Sex Research*, 2002).*

KYUNG-EUN PARK has studied adopted adolescents, especially their knowledge about and contact with their birth parents, for the past five years. She is working at Gyeonggido Family and Women Development Institute in Korea as a researcher for developing family policy. Her recent research topics are family-friendly policy and low-income families. She has published "Marriages and Families in the United States" (with B. C. Miller, S. C. Leavitt, and M. Junius, in Families Global Perspective*, Allyn & Bacon, 2004) and "Adoption Outcomes" (with B. C. Miller and B. W. Winward, in* Historical and Cross-Cultural Encyclopedia of Adoption*, Greenwood, in press).*

BRYAN W. WINWARD has recently completed his PhD at Utah State University in marriage and family relations. His research has focused on comparing adopted and nonadopted young adults on the type and quality of their intimate relationships. His work includes "Treatment and Prevention of Adolescent Pregnancy" (with B. C. Miller and R. Sage, in The Handbook of Dysfunctional Behavior in Adolescence: Theory, Practice, and Prevention*, Kluwer, in press) and "Adoption Outcomes" (with B. C. Miller, B. W. Winward, and K. E. Park, in* Historical and Cross-Cultural Encyclopedia of Adoption*, Greenwood, in press).*

MATHEW CHRISTENSEN works as a statistical analyst and epidemiologist at the Colorado Department of Public Health and Environment in Denver, Colorado. He completed his PhD at Utah State University, specializing in human/adolescent development and social research. In his current position, he consults and provides technical support about human development and public health through the use of statistics, survey design, data collection, and program evaluation. Publications include "Questionnaire and Interview Inconsistencies: Exaggerated Differences between Adopted and Nonadopted Adolescents in a National Sample" (with X. Fan, B. C. Miller, K. Park,

H. D. Grotevant, M. van Dulmen, N. Dunbar, and B. K. Bayley, Adoption Quarterly, *2002) and "Comparisons of Adopted and Non-adopted Adolescents in a Large, Nationally Representative Sample" (with B. C. Miller, X. Fan, H. D. Grotevant, and M. van Dulmen,* Child Development, *2000).*

HAROLD D. GROTEVANT is a distinguished university teaching professor of family social science and adjunct professor of child psychology at the University of Minnesota. His research focuses on relationships in adoptive families and on identity development in adolescents and young adults. His work has resulted in more than one hundred articles published in professional journals as well as several books. Publications include "Antisocial Behavior of Adoptees and Nonadoptees: Prediction from Early History and Adolescent Relationships" (with M. H. van Dulmen, N. Dunbar, J. Nelson-Christinedaughter, M. Christensen, X. Fan, and B. C. Miller, Journal of Research on Adolescence, *2006) and "Openness in Adoption: Outcomes for Adolescents within Their Adoptive Kinship Networks" (with Y. Perry and R. G. McRoy, in* Psychological Issues in Adoption: Research and Practice, *Praeger, 2005).*

ROBERT H. TAI is an assistant professor of science education in the Curry School of Education, University of Virginia. He is currently a co-principal investigator of a large-scale national survey study using college students' self-reports of their high school science experiences as predictors of introductory college science performance. Recent publications include "Factors Influencing Success in Introductory College Chemistry" (with P. M. Sadler and J. F. Loehr, Journal of Research in Science Teaching, *2005) and "The Help-Seeking Orientations of Latino and Non-Latino Urban High School Students: A Critical-Sociological Investigation" (with R. D. Stanton-Salazar and L. F. Chávez,* Social Psychology of Education, *2001).*