An Experimental Test of the Accuracy of Proxy Reports with External Validation

A Thesis
Submitted to the Program in Quantitative Methods for Social Sciences
of Columbia University
in partial fulfillment of the requirements
for the degree of
Master of Arts

Curtiss L. Cobb III
September, 2009

This work was carried out under the guidance and support of Jon Krosnick, Frederic O. Glover Professor in Humanities and Social Sciences, Departments of Communication, Stanford University.

Correspondence about this manuscript should be addressed to Curtiss Cobb, 450 Serra Mall, 120-160, Stanford University, Stanford, CA 94305 (clc2003@stanford.edu)

Experimental Test of the Accuracy of Proxy Reports with External Validation

Abstract

Survey researchers routinely use responses from proxies when information is hard to collect from targets, which raises questions about how accurate proxy reports are relative to self-reports. There is a large body of research that tries to ascertain the quality of proxy responses; however, the designs of many of these studies make it difficult to draw any generalizable inferences with confidence. We report the results of two studies: (1) the accuracy of alumni from Stanford University and the University of Virginia reporting the field(s) of their bachelor's degree in omnibus alumni association telephone surveys; and (2) the accuracy of Stanford alumni and a parent reporting the alumnus's field(s) of bachelor's degree using a two-question Internet survey. We found that both alumni and parent-proxies are extremely accurate in identifying fields of degree in an open-question format, though alumni out-perform their parents by 10%. Alumni and parents are more accurate when answering a closed question with 7 response options rather than 10 response options. Agreement for alumni and parent errors is high, and alumni errors on the closed question are a greater predictor of parent closed-question errors than parent open-question errors.

An Experimental Test of the Accuracy of Proxy Reports with External Validation

The use of proxy respondents is a common practice in survey research. Proxy responses make up a significant proportion of survey interviews, including surveys known to be of high quality. For example, both the U.S. Census and the Current Population Survey ask one member questions about an entire household rather than attempting to interview each household member separately. By some calculations, proxy responses account for 50% or more of the conducted interviews for the U.S. Census Bureau's surveys (Boehm 1989). Studies that have used proxy responses span a variety of subjects, including topics as diverse as health and illness, social stratification and income, immigrant issues, employment, political participation, and others.

The use of proxy respondents in survey samples has many practical advantages that make it appealing to survey researchers. Most notably, using proxy respondents alongside target respondents can make obtaining information easier and cheaper. Proxies can increase contact rates and cooperation rates when the targets themselves are hard to interview. For example, surveys of immigrant populations have used proxies when language barriers may otherwise hinder participation. Proxies can reduce costs by providing information about multiple targets in a single interview. By some estimates, proxy responses lowers the total cost of the Current Population Survey by up to 17% compared to what it would cost to actually interview each target (Boehm 1989). Additionally, there are situations in which some researchers believe that interviewing proxy respondents over target respondents yields better information, such as when social desirability effects are believed to be possible or when parents report on young children.

The most obvious concern about using proxy respondents is how accurate proxies are in providing second-hand information about target respondents. Is the information provided by proxies less accurate than self-reports, or biased in systematic ways that affect the conclusions researchers draw from the data? In fact, such questions have spurred a large body of research extending over more than fifty years that has sought to understand the relationship between target responses and proxy responses and its effect on survey quality. However, the results of these studies vary considerably, both between and within subject domains and across question types, leaving many scratching their heads as to when, or even whether, proxy responses are "safe" to use. In fact, one of the most consistent findings within this body of work is how inconsistently "accurate" proxy responses tend to be (Menon et al. 1995; Moore 1988; Mosely and Wolinsky 1986; Looker 1989).

In this paper, we explore the accuracy of proxy reports relative to target self-reports by asking college graduates and their parents about the graduate's field(s) of bachelor's degree. We validate the accuracy of responses externally using official records provided to us by the universities that the graduates attended. Additionally, we examine how proxy reporting errors are related to target reporting errors in terms of both magnitude and direction. We begin by outlining the theoretical ways that proxy respondents may differ from target respondents in how they answer questions, which may lead to lower accuracy for proxy responses. Then, we review the results of past studies that at first seem to offer evidence regarding these hypotheses. But as we will explain, the designs of these studies render most of them uninformative about the relative accuracy of proxy reports. Next, we briefly examine the accuracy of self-reports using data from two telephone surveys of alumni from Stanford University and the University of Virginia. Last, we describe the results of an experiment with a matched sample of alumni from Stanford University and their parents in which both groups were asked the same questions.

LITERATURE REVIEW AND HYPOTHESES

The general process by which survey respondents, whether targets or proxies, answer questions is often modeled simplistically in four parts: (1) Comprehension—the respondent interprets the meaning and intent of each question; (2) Retrieval—the respondent recalls the information necessary to answer the question; (3) Judgment—the information is summarized into a response to the question asked; and (4) Reporting—the respondent maps his/her answer onto the given response format for the question (Groves, Fowler & Couper 2004; Krosnick 1991, 1994). To answer survey questions optimally (with the most accuracy), each of these steps must be carried out completely and in sequence. Many respondents, whether targets or proxies, may initially be willing to exert the effort necessary to answer questions optimally prior to starting the survey, but later find themselves unable or unwilling to do so. There is strong evidence that when respondents incompletely carry out any one of these processes, the result is measurement error (Krosnick 1999). This behavior, known in the literature as satisficing, is more likely to occur when survey questions are difficult and place heavy demands on memory or require complicated judgments. It is possible that the same question could place different cognitive demands on targets and proxies, leading to different levels of response quality.

Target and proxy responses can vary in their understanding of the meaning and intent of the questions they are being asked, particularly when the questions are specific to some aspect of the target and not asked of a general population. For example, targets and proxies may interpret questions on medical questionnaires that are specific to an ailment that the target suffers from differently because of their different exposure to medical professionals treating the ailment. Targets and proxies can also differ in their ability to retrieve the information necessary to answer questions about the target. Most simply, the target and/or proxy may not know the information necessary to answer the question(s) being asked about the target, while the other individual does have the information. Further research has shown that autobiographical information is associated with enhanced recall relative to other information not self-referenced, which makes the retrieval process easier for a target respondent than a proxy respondent (Bower and Gilligan 1979; Klein et al. 1989). Enhanced recall may result in higher accuracy or over-reporting.

Target and proxy respondents may also differ in the extent to which they can consolidate the known information into an accurate summary judgment. They can differentially weigh available information even when the same information is available to both the target and the proxy. Mapping judgments onto given response options requires the respondent to understand the distinctions between response options. A number of studies related to children's reports of parents SES characteristics find that children's answers become more similar to their parents answers as the children become older up until the time the child leaves the home, then drop the longer the child lives outside the home (Cohen & Orum 1972; Looker 1989). Looker (1989) hypothesizes a socialization mechanism through which parental characteristics will be accurate as long as the information is salient to the child and it is something they are likely to know. The ability to map judgments onto given response options, therefore, should vary for targets and proxies based on differences in their familiarity with the topic of the survey.

To summarize, questions that require less cognitive effort should result in more accurate responses than questions that require greater cognitive effort for both targets and proxies. In cases where the cognitive demand necessary to answer a question accurately is different for the target and the proxy, then the accuracy of their responses should differ. If the cognitive demand is greater for the target than for the

proxy, then the target should have lower rates of response accuracy. If the cognitive demand for the proxy is lower than that of the target, then the proxy should respond with less accuracy.

At first glance, many past studies appear to be useful for testing under what conditions and in what ways targets and proxies will vary in response accuracy to survey questions because these studies compare answers from proxy reports to answers from target reports. However, upon close inspection, the designs of most of these studies make it difficult to draw any generalizable inferences with confidence. We now explain the design features that studies must have in order to be informative about the accuracy of proxy reports relative to target self-reports. Then, we describe the studies that do and do not meet these criteria.

Necessary features of a study. In order to identify the differences in the accuracy rates of target responses and proxy responses in surveys clearly, a study should have the following four characteristics:

First, both target respondents and proxy respondents should be interviewed. Second, target respondents and the targets that proxies are asked about should both be representative samples of the same population. This can be fulfilled either by randomly assigning targets to respond for themselves or have a proxy respond for them or by interviewing matched pairs of targets and proxies. Only studies with matched target/proxy pairs can reveal the relationships between target reporting errors and proxy errors. Some studies that compare target/proxy responses have tried to obtain information first from a target respondent and then from the proxy only if the target was unavailable. These studies tend to conflate the independent effects of reporting bias, or bias in the answers given, and sampling bias, bias in who is giving answers, treating everything as reporting bias.

Third, the questions used to compare accuracy should be asked identically in the two sets of interviews, and they should be asked in identical contexts; that is, the number, content, sequence and modes of the questions being asked should all be the same. Finally, there should be an independent external measure of accuracy by which responses are evaluated[1]. The majority of studies evaluating the accuracy of proxy reports confound target/proxy agreement with accuracy, yet targets and proxies can both be wrong and still agree.

Identifying useful studies. After conducting an exhaustive literature search, we uncovered 79 studies that compared proxy reports to target reports; these studies are listed down the left side of table 1. Some of these studies are potentially useful for our purposes, but many of them are not. The *X*'s in each row of the table indicate the features of that study that render it uninformative with regard to the accuracy of proxies relative to target respondents.

The first column of *X*'s identifies studies that interviewed proxy respondents only. These studies often compared proxy response distributions to target responses from another study. The second column identifies studies that obtained proxy reports only when targets were unavailable. Column 3 shows which surveys used different questionnaires, modes, or instruments to obtain responses. Column 4 identifies the large number of studies that rely on target/proxy agreement as a measure of proxy accuracy.

[1] Measures of accuracy are only possible for survey questions measuring objective phenomenon. There is no direct way to know about people's subjective states independent from what they tell us. Subjective phenomena can only be evaluated based upon consistency between target and proxy responses. The source and direction of reporting error cannot be ascertained.

Six studies do not have any of the limitations noted in Table 1, and they all involved reports of medical events. In one study (Cobb et al. 1956), self-reports of arthritis or rheumatism diagnoses were accurate for 60% of respondents, and proxy reports were accurate only slightly less often: 56%. Thompson and Tauber (1957) found that self-reports and proxy reports of whether people had received a diagnosis of heart disease were accurate equally often: 73% for both self-reports and for proxies. Andersen et al. (1973) reported that targets and proxies were equally accurate when reporting physician visits (targets 55% versus proxies 54%); Balamuth (1965) reported similar results for hospitalizations (88% versus 87%). Parents were more accurate in reporting children's doctor visits than children (Cannell and Fowler 1963).

Only one study provided different results: Magaziner et al. (1997) examined data on the performance of daily physical and instrumental activities among people who had broken their hips. They found that proxy responses about physical activities deviated from self-reports by an average of 18% and from direct observations by 41%; these deviations were 3% and 37%, respectively, for instrumental activities.[2] Unfortunately, Magaziner et al. (1997) do not provide statistical analyses of the differences between self-report and direct observations. However, the bias between proxies versus self-reports and proxies versus direct observations are in the same direction, suggesting targets are more accurate than proxies.

Thus, most studies without design flaws found equal accuracy for self-reports and proxy reports, though one study found proxy reports to be notably less accurate than self-reports. On these grounds, one might view this literature as suggesting that with isolated exceptions, proxy reports may yield equivalent accuracy to self-reports; thus, there would be no reason to prefer self-reports over proxy reports.

To see whether this conclusion holds outside of the domain of medical care, our study examines student experiences in college. Specifically, we present the results of two studies examining how accurate responses are to questions about the field(s) of targets' bachelor's degree(s). The first study examines the accuracy of target responses to a telephone survey using alumni from Stanford University and the University of Virginia (UVA). The second study compares the accuracy of target and proxy responses using an Internet survey of Stanford University alumni and their parents. We validate responses for both studies using official records from the registrar's office of the target's alma mater. This work is part of a larger set of studies commissioned by the National Science Foundation (NSF) to optimize question wording for a field(s) of degree question to be added to the American Community Survey (ACS) to identify all individuals with science and engineering training, as defined by the NSF, in order to draw a representative sample for NSF-administered surveys. The ACS, like other Census Bureau surveys, relies heavily on proxy respondents to report information for other household members.

The purpose of the first study is to understand how different question types and response options impact the ability of targets to identify the field(s) of their bachelor's degree correctly. The ability to survey alumni from two schools gives us a more generalizable understanding of how different factors relate to accurate responses. The second study builds on the first by adding a sample of matched alumni and their parents, which allows us to see if there are different accuracy rates for targets and proxies, as well as whether target accuracy and proxy accuracy are related to one another. In essence, when one makes an error, how likely is the other to make the same error? Given that much concern lies

[2] Magaziner et al. (1997) report Cohen's kappa statistic, a measure of agreement. The average kappa for proxy versus direct observation is .38, which suggests little agreement. The average kappa for proxy versus self-reports is .49, which suggests moderate agreement.

in the relative accuracy of proxies to targets, we are able to see how much proxy-reporting error is shared by the target.

It is difficult to hypothesize with any certainty whether the alumni or their parents will be more accurate at correctly identifying the alumni's field(s) of degree. We could predict that we are likely to find no difference between targets and proxies given the results of five of the six studies described above. This is particularly true if the proxy is merely relaying information provided to him/her by the target (whereby target accuracy would present an upper-bound on proxy accuracy). On the other hand, we can adopt the prevalent assumption found in most of the other studies purporting to compare target responses to proxy responses—that when a target is fully mature and cognitively capable, he/she should be more accurate than a proxy. One can even imagine a scenario in which a parent learns of the alumnus's major(s) independent of the alumnus; with less overall information about the alumnus's college experience and coursework to sort out cognitively during the retrieval process, the parent may be able to make a more accurate report of the alumnus's major(s). Therefore, we have designed our study to be able to test whether target and proxy responses differ from each other in terms of accuracy and, if so, by what magnitude and in which direction. We also test whether the matched alumni and parents are making the same types of errors (which would suggest that source of the error is the target and not the proxy).

Both studies ask all respondents to answer the same two questions related to the alumni's field(s) of degree: (1) a closed-ended question to categorize the field(s) of degree; and (2) an open response question asking respondents to state the alumni's specific field(s) of degree. Both open and closed questions can lead to misreports wherein the respondent fails to accurately recall the information necessary to the answer the survey question. However, closed questions are also susceptible to misclassification, wherein respondents know the correct answer but are unable to map it onto the predefined set of response options in the way the designer of the question wishes. Open response questions have the benefit of omitting the need to map answers onto response options and usually yield higher quality responses. Nevertheless, coding open responses generally takes considerable manpower and financial resources to complete and is very time consuming, making it impractical for many survey researchers. Therefore, optimizing the design of a closed-ended question has considerable appeal if the question can be designed to yield accurate responses. Alternatively, closed response options, if sufficiently broad, may make it easier for a respondent who does not know specific or complete information to be more likely to give an accurate response. In our study, we are able to distinguish whether target and proxy errors are misreports (using the open question) or misclassifications (using the closed-ended question) and test their prevalence and relationship to each other and the type of respondent.

Given the preference many survey researchers have for closed-ended questions, we asked the closed-ended field(s) of degree question using seven response categories and ten response categories. There are many possible logical explanations for how the number of field(s) categories proffered in the field(s) of degree question might affect the cognitive requirements of classifying one's degree. Fewer categories might be overly broad and vague, making classification difficult. A larger and more specific list might make it possible for the respondent to find a category that is a closer match to his/her field(s) of study and increase confidence in her classification. On the other hand, an overly specific list may increase difficulty of classification by requiring respondents to understand fine distinctions between categories. An overly specified list may lead to over-classification, whereby respondents answer affirmative to too many categories, or under-classification, where a respondent does not answer affirmative to enough categories. This may be especially the case given the rise of multidisciplinary

majors that span categories. Additionally, there may be some social desirability in over-reporting majors, which would affect both the closed and open questions. We test whether seven or ten response categories yields more accurate data for targets and proxies, as well as examine relative rates of over-classification and under-classification.

Finally, we control for a number of alumni and parental factors that may influence reporting accuracy. First is the year the alumni graduated with their bachelor's degrees. The saliency of the field(s) of bachelor's degree may fade over time in the minds of both alumni and their parents. Additionally, parents with a personal college experience may understand the fields of degree and the differences between them more than parents that did not go to college, making the question easier to answer. Unfortunately, we do not have a measure of parental education level. Instead, we use parent's title as a proxy and identify those parents whose title suggests they hold an advanced degree. A number of studies offer insight into the associations between the characteristics of the target and/or the proxy and how they relate to agreement between the two. Many studies found that women report more symptoms and incidences on medical surveys for both themselves and others (Andresen et al. 2001; Bassett et al. 1990; Briscoe 1984; Chen et al. 2007; Clarridge & Massagli 1989). Therefore, we control for both the gender of the parent and the gender of the alumnus.

Next, we discuss the results of our two studies.

STUDY 1: ACCURACY OF ALUMNI FROM STANFORD UNIVERSITY AND UNIVERSITY OF VIRGINIA WHEN RESPONDING TO QUESTIONS ABOUT THEIR BACHELOR'S FIELD(S) OF DEGREE.

Methods. We conducted two separate telephone surveys of alumni from Stanford University and the University of Virginia (UVA) to assess the accuracy of target respondents in classifying the field(s) of their bachelor's degrees. The procedures and question wording for the Stanford and UVA surveys were identical. Details about the sampling procedure, response rates and question parameters for each school are as follows:

Both telephone samples consisted of undergraduate alumni living within the continental USA or Canada who earned their undergraduate degrees from 1955 to 2005; individuals were contacted between May 19 and June 2, 2006. Interviewing quotas were established to obtain proportional representation by class year (so that older alumni, who are most likely to be at home in the evening, were not over-represented). Quotas were also established within each class year grouping to obtain proportional representation by donor status (so that donors, who are more likely than non-donors to agree to be interviewed, were not over-represented). The samples were drawn to be proportionally representative of the undergraduate alumni population by all demographic measures including class year, donor status, gender, school, degree level (undergraduate-only or dual), alumni association membership, geographic region, and ethnicity.

Attempts were made to contact 5,508 eligible Stanford alumni and 5,313 eligible UVA alumni using the interviewing firm Telepoll Canada. At least eight call attempts were made before a number was retired from the sample. In the end, 1,658 Stanford alumni were successfully contacted (contact rate of 30.1%), with 634 completed interviews (11.5% response rate and 38% cooperation rate); 1,742 UVA alumni

were successfully contacted (contact rate of 32.8%), with 573 completed interviews (10.8% response rate and 32.9% cooperation rate).

While the surveys covered many topics, included was a test of how accurately all graduates were able to report the field(s) of study of their bachelor's degrees using one of four question response formats originally proposed by the National Science Foundation. All respondents were asked the following question:

> *"Thinking only about the bachelor's degrees you received from [Stanford/UVA], which of the following broad categories best represent the field or fields of your major(s)? Did you get a bachelor's degree in…*

Respondents were randomly assigned to one of the following four response conditions: (1) 7 field categories, original order; (2) 10 field categories, original order; (3) 7 field categories, reverse order; and (4) 10 field categories, reverse order. Respondents were asked to provide an affirmative or negative response ("Yes/No") for each category listed. The response categories used were:

<u>7 Categories, Original Order</u>

A. Biological, agricultural, physical, or related sciences
B. Health, nursing, or medical fields
C. Engineering, computer, mathematical, or related sciences
D. Psychology, economics, or other social sciences, except history
E. History, arts or humanities
F. Business, communication, or education
G. Some other field(s)

<u>10 Categories, Original Order</u>

A. Biological, agricultural, or related sciences
B. Health, nursing, or medical fields
C. Computer, mathematical, or related sciences
D. Engineering
E. Physical or related sciences, including earth sciences
F. Psychology, economics, or other social sciences, except history
G. Business or communication
H. Education
I. History, arts or humanities
J. Some other field(s)

The ten response categories were nested within the seven response categories. For the reverse order conditions, the "Some other field(s)" category (G and J, respectively) was always read last.

Both Stanford and UVA provided us with data on respondents' specific majors from their official records to measure the accuracy of respondents' answers. The reported majors provided by the universities were classified into the seven and ten category response options using a cross-walk provided by the National Science Foundation. Specific majors, particularly interdisciplinary/multidisciplinary majors, can map onto more than one response category.

All respondents were also asked the following open-ended question regarding their field(s) of degree immediately after the question using the list of field(s) categories:

> *"In what specific field(s) or fields did you receive a bachelor's degree from [Stanford/UVA]? _____"*

Table 2 reports the univariate statistics separately for both the Stanford and University of Virginia samples.

Results. Responses to the open-ended question were overwhelmingly accurate. A total of 92.9% of Stanford respondents and 83.4% of UVA respondents correctly named the major(s) of their bachelor's degree. An additional 4.7% and 12.6%, respectively, were partially correct in naming their majors (with either majors incorrectly added or incorrectly omitted). Only 2.4% of Stanford respondents and 4.0% of UVA respondents were completely wrong when naming their major(s) in the open-ended question. For both Stanford and UVA, respondents in the seven-category, original order closed-question condition reported completely wrong majors in the open question more than respondents in the other closed response conditions (4.1% for Stanford and 4.6% for UVA, see Table 3). Using the open question to identify those with science and engineering training was very successful across all response conditions, with accuracy rates of 95.7% or higher. In fact, the number of science and engineering classification errors made was so few that regression analysis was not possible.

A sizeable majority of respondents to both school surveys correctly classified the major(s) of their bachelor's degree using the closed-ended question. Stanford alumni correctly classified their field(s) of degree 69.4% of the time, while alumni at UVA did so 65% of the time. Another 14.8% of Stanford respondents and 17.3% of UVA respondents chose the correct category as well as additional erroneous categories. Only 11.5% and 17.7%, respectively, completely misclassified their field(s) of degree. In addition, 4.3% of Stanford respondents failed to answer affirmative to all of the necessary categories while correctly answering affirmative to some of the categories. These respondents were reporting on interdisciplinary degree programs, which were not present in the UVA sample.

The *7 Category, Original Order* response condition yielded the most accurate results for both schools (71.8% for Stanford and 69.7% for UVA; see Table 4). It also had the lowest misclassification rates for both samples, 9.4% and 15.9% respectively. The *7 Category, Reverse Order* condition yielded the lowest proportion of correct classifications (66.5%) and the highest proportion of misclassifications (13.3%) for the Stanford sample. The *10 Category, Reverse Order* condition had the lowest levels of correct classifications (58.7%) for the UVA sample, while the *10 Category, Original Order* condition yielded the highest proportion of misclassifications (19.2%).

The rates of accuracy are higher when using the categories to identify potential respondents for a Science and Engineering sample.[3] Among Stanford respondents, 91.8% correctly classified themselves into a Science and Engineering category; 85.6% of UVA respondents did so. The *7 Category, Original Order* response condition yielded the highest accuracy for both schools (96.5% and 88.6%, respectively; see Table 2). Also consistent between the schools was that the *7 Category, Reverse Order* yielded the highest proportion of respondents that misclassified themselves out of Science and Engineering when

[3] NSF is particularly interested in using a field of degree question in its surveys to identify a sample of respondents with degrees in science and engineering fields.

they should have been included (8.9% and 11.9%, respectively). The *7 Category, Original Order* had the highest rate of respondents misclassified into Science and Engineering for Stanford (3.8%), while the *10 Category, Original Order* condition had the highest rate for UVA (5.4%).

The strongest predictor of making a closed-ended question classification error is making an open-ended response error (see Table 5). Misreporting in the open question yields an increase in the log-odds of committing a closed-question classification error of b=1.47 (p<.01) for Stanford respondents and b=1.69 (p<.001) for UVA students. Over-reporting or under-reporting in the open question is also strongly related to making closed-ended classification errors (Stanford over-report b=1.51, p<.001; UVA over-report b=1.07, p<.01; UVA under-report b=1.28, p<.01). Class year is a marginally significant predictor of making complete misclassification errors. For each year later that a respondent graduated, the log-odds of a classification error are increased by b=.01 (p<.1). Response condition was only significant in one case: the *10 Category, Reverse Order* condition is associated with higher classification errors for UVA respondents (b=.54, p<.05). Finally, open-question errors are the strongest predictor of making Science and Engineering classification errors on the closed-ended question (see Table 5). Total misreport increased the log-odds for UVA respondents (b=1.15, p<.05), while over-reporting in the open question increased the log-odds for Stanford respondents by b=2.05 (p<.001). The reverse order response conditions increase the log-odds of Science and Engineering classification error for Stanford respondents (7-category b=1.52, p<.01; 10-category b=1.20, p<.05), but not for UVA respondents.

Conclusion. The results of the telephone alumni studies presented here reveal that much of the error in responding to closed questions is a result of not having the correct information to report rather than from the task of mapping the responses onto the provided response categories. We did not find significant differences in accuracy between the four question versions we examined; however, the *7 Category, Original Order* condition performed best. The results of these studies are generally reassuring given their high rates of accuracy. Of course, it would be ideal if no respondents made mistakes, but improving upon these accuracy rates may be especially challenging.

STUDY 2: ACCURACY OF ALUMNI FROM STANFORD UNIVERSITY AND THEIR PARENTS IN RESPONDING TO QUESTIONS ABOUT THE ALUMNI'S BACHELOR'S FIELD(S) OF DEGREE.

Methods. We conducted a similar two-question Internet survey of field(s) of degree with Stanford alumni graduating in 1970 or later and a matched parent. Alumni and their parents were asked modified versions of the same open and closed questions used in the Stanford and UVA telephone surveys. The closed question for alumni read:

> "*Thinking only about the bachelor's degree(s) you received from Stanford, which of the following broad categories best represents the field(s) of your major(s)? Did you get a bachelor's degree in..."*

While the closed question for parents read:

> "*Thinking only about the bachelor's degree(s) [child's name] received from Stanford, which of the following broad categories best represent the field(s) of [child's name]'s major(s)? Did [child's name] get a bachelor's degree in..."*

Alumni and their matched parents were assigned to the same seven category or ten category response versions and were asked to provide an affirmative or negative response ("Yes/No") for each category listed. The categories were identical to the original order categories used in the Stanford and UVA telephone surveys.

The open-ended follow-up question for alumni read:

"In what specific field(s) did you receive a bachelor's degree?"

The open-ended follow-up question for parents read:

"In what specific field(s) did [child's name] receive a bachelor's degree?"

Both alumni and parents were contacted about the survey via email. A pre-notice was sent one week before the e-mailed invitation containing the survey link in order to alert respondents to the survey. Up to two follow-up reminder e-mails were sent to those Stanford alumni and parents that did not respond to the initial invitation. Parents were surveyed between November 8, 2006 and January 7, 2007, while alumni were surveyed between March 9, 2008 and April 6, 2008. The sample consisted of 4,021 alumni and 3,653 parents, with 368 parents being asked about two alumni children. A total of 3,713 e-mail invitations were successfully sent to the alumni and 2,013 questionnaires were completed (50.1% response rate and 54.2% cooperation rate); 1,724 matched parents completed questionnaires for 1,893 alumni (47.2% response rate). Table 6 reports the univariate statistics of both alumni and parent respondents.

Results. The open-ended question asking respondents to name the alumni's field(s) of degree specifically yielded a high rate of accuracy: 96.7% of alumni and 88.7% of parents correctly named the major(s) of the alumni's bachelor's degree. An additional 2.8% and 4.7%, respectively, were partially correct in naming the majors (with either majors incorrectly added or incorrectly omitted). The most common error made by alumni was to over-report majors (2.3%), while parents' most common mistake was to completely misreport (6.6%). The 10-point difference in accuracy rates between alumni and parents diminishes to less than 2% when looking at how well respondents' answers identify them as science and engineering, with alumni accurate 99.4% of the time and parents accurate 97.6% of the time. There was no difference in open response accuracy by closed-ended response conditions (see Table 7).

The accuracy rates of alumni responding to closed-ended questions and parents reporting on alumni field(s) of degree look very similar. The alumni respondents correctly classify their field(s) of degree 69.8% of the time in the *7 Category* response condition (compared to 64.8% for parents; see Table 8) and 58.3% of the time in the *10 Category* response condition (compared to 56.2% for parents). For both alumni and parents, the difference between the response conditions is statistically significant when looked at alone (Alumni $X^2(3)=31.72$, $p<.001$; Parent $X^2(3)=16.22$, $p<.001$).

When alumni and parents make classification errors, their errors look different. Alumni are more likely to make additional erroneous classifications along with a correct classification (*7 Category* 14.1% and *10 Category* 22.1%; see Table 8), while parents are more likely to make completely erroneous classifications (*7 Category* 18.2% and *10 Category* 24.9%). The difference in the distribution of errors between alumni and parents is significant within the response category conditions (*7 Categories* $X^2(3)=19.76$, $p<.001$; *10*

Categories $X^2(3)=53.86$, $p<.001$). The open-ended follow-up question asking respondents to name the alumni's field(s) of degree specifically yielded much more accurate results than the closed-ended question; 96.7% of alumni and 88.7% of parents correctly named the major(s) of the alumni's bachelor's degree. An additional 2.8% and 4.7%, respectively, were partially correct in naming the majors (with either majors incorrectly added or incorrectly omitted).

The strongest predictor of making a closed-ended question classification error for both alumni and parents is making an open-ended response error (see Table 9). A misreport in the open question yields an increase in the log-odds of committing a closed-question classification error of b=1.57 (p<.01) for alumni respondents and b=2.00 (p<.001) for parents. Over-reporting or under-reporting in the open question is also strongly related to making closed-ended classification errors (Alumni over-report b=1.53, p<.001; Parents over-report b=1.35, p<.001; Parents under-report b=.89, p<.05). Class year is a significant predictor of making classification errors for alumni (b=.03, p<.001), but not for their parents (b=.01, n.s.). Response condition was a significant predictor for both parents and alumni, with *10 Categories* yielding more classification errors (Alumni b=.50, p<.001; Parents b=.34, p<.01) than *7 Categories*. Parents answering for female alumnae are more likely to make a classification error than parents answering for males (b=.29, p<.01). The gender of the parent or the number of children that the parent was asked to report about is not related to the likelihood of making an error.

There is no significant difference in the accuracy rates or types of errors for parents whose children answered the alumni questionnaire versus those whose children did not answer the alumni questionnaire (closed X^2 (3)=0.85, n.s.; open $X^2(3)$= 7.39, n.s; see Table 10). Parents whose children responded to the survey correctly reported 60.7% in the closed-question and 89.3% in the open question, while those whose children did not reply to the survey correctly reported 59.9% for the closed-question and 88.1% in the open question.

When we restrict our sample to just those alumni whose parents also answered the survey, we find that the accuracy of alumni and parents, as well as the types of errors they make, are closely related (see Tables 11 & 12). Ninety percent of the parents whose children answered the open-ended question correctly answered it correctly themselves. The majority of parents also reported correctly when their children over-reported (68.4%) in the open question. Of alumni who correctly classified their field(s) of degree using the closed-question, almost 79% of their parents also were correct, while 58.8% of the alumni who misclassified also had parents that misclassified. On the other hand, 69.4% of alumni had parents that incompletely classified when they also incompletely classified. The only type of error made by alumni in the closed question that was not also likely to be matched by their parents was over-classifying their field(s) of degree. In those instances, parents were most likely to answer correctly (44.6% compared to 20.4%).

We also evaluated how related alumni and parent responses are in terms of how both answered each category of the closed-ended question. The average percent agreement is quite high, 91.3% for the *7 Category* condition and 91.9% for the *10 Category* condition (see Table 13).[4] For those whose parents make any type of error, the percent of agreement is still above 80% for both conditions; when parents

[4] We report percent agreement and Cohen's kappa statistic (see Table 13). The kappa statistic allows us to capture the proportion of agreement beyond what is attributable to chance alone. The guidelines for interpreting the kappa statistic are as follows: 1 is perfect agreement, >.8 is almost perfect agreement, 0.6-0.8 is substantial agreement, 0.4-0.6 is moderate agreement, and <0.4 is slight to fair agreement (Landis & Koch 1977).

make a misclassification error, the percent of agreement is greater than 75% for both conditions. The category of *Psychology, economics, or other social sciences, except history* yielded the least percent of agreement for both the *7 Category* and *10 Category* conditions, including the cases in which parents made an error. When comparing all alumni-parent matches, two categories from the *7 Category* condition and three from the *10 Category* condition result in small kappas, which means little agreement by chance. Interestingly, these categories are in fields that are not offered to undergraduates by Stanford, such as education or health or the "some other field(s)" category. The social sciences and humanities in both conditions and the physical sciences in the *10 Category* condition result in small kappas as well when we look only at agreement between parents making any error and their matched children. The results are similar for parents making only a classification error.

The strongest predictor of matched parents making a closed-ended question classification error is the alumnus child making a closed-ended classification error (see Table 14). Misclassification by children increases the log-odds of a parent making any classification error by b=2.88 (p<.001) and making a misclassification error by b=2.70 (p<.001). These coefficients are larger than even when parents make their own open-question reporting errors. Parents who misreport in the open question increase their log-odds of making any error by b=1.91 (p<.001) and of making a misclassification error by b=1.63 (p<.001). The same pattern is found when we examine whether parents' responses can be used to identify alumni with science and engineering training. Alumni closed-question errors increase parents' Science and Engineering errors by b=1.94 (p<.001), and parents' open errors increase the log-odds by b=1.61 (p<.001). Over-reporting and under-reporting by alumni are also strongly related to parents making an error. If alumni over-classify, then parents' log-odds increase by b=1.64 (p<.001) for making any error and b=1.43 (p<.001) for a misclassification. Under-classification by alumni relates to increases of b=5.24 (p<.001) and b=1.02 (p<.001), respectively.

Parents of more recent graduates make fewer Science and Engineering errors (b=-.04, p<.01), while parents of female alumni make more Science and Engineering errors (b=.46, p<.01). Parents with advanced degrees are less likely to make a misclassification error (b=-.42, p<.1) and less likely to make a Science and Engineering error (b=-.64, p<.05).

CONCLUSION

The results of these studies are generally reassuring. Both alumni and parent-proxies are extremely accurate in identifying fields of degree in an open-question format and do quite well when asked to classify the degrees according to closed-question response options. While alumni are almost 10% more accurate in the open-question format, parents, as proxies, perform almost as well as alumni in the closed question format (difference of less than 5%). In fact, the accuracy of alumni and parents within the same closed-ended response category conditions is more similar than scores between alumni in different conditions. Even though the difference in accuracy is slight, it is still statistically significant, which somewhat supports suspicions that other researchers have about the quality of proxy data.
However, when errors are made, alumni and parents make very similar errors. Alumni errors are a stronger predictor of the parent making a closed-question classification error than the parent making an open-question error. This is highly suggestive that parents got their information from the alumni and corroborates Magaziner et al.'s (1997) finding that proxy agreement with targets was greater than proxy agreement with external medical evaluations.

Proxy responses are generally thought by others to be biased in the direction that they deviate from target responses. There are, of course, exceptions. When strong social desirability effects are believed to be possible or when adults are responding for children, the opposite assumption is often made. In our study, where there is a difference in the types of errors made by alumni and parents, the results suggest that parent errors are less biased than alumni errors. While parents get the classifications entirely wrong, alumni are more likely to over-classify their field(s) of study using the closed question. Frequently, studies claim that proxies under-report because they do not report the same rates of incidences at targets. Rather, we find that targets are over-reporting their field(s) of degree.

The types of response options used by the survey certainly make a difference. Across both studies and all conditions, the open question far out-performed the closed question. The difference in accuracy between the open and closed questions, which averaged almost 30 points, suggests that much of the problem for both alumni and parents is not identifying the major(s), but classifying them according to the closed response options provided. Additionally, seven response categories yielded more accurate results than ten response categories.

Parents known to have advanced degrees were more accurate than parents without advanced degrees. An advanced degree can serve as a proxy measure for how familiar the parent might be with academic institutions and the disciplinary boundaries that exist therein. This finding supports the notion that proxy respondents for whom the information is more likely to be salient are better equipped to report accurately on the target than proxies unfamiliar with topic of interest. Being asked about multiple children did not lead to any more error than being asked about one child. The gender of the parent and the gender of the alumnus were marginally significant predictors of a respondent's ability to correctly classify an alumnus's degree.

The data analyzed here were collected from alumni and parents of alumni at two relatively elite universities, and the vast majority of American college students receive degrees from less elite schools. The parents in our sample also had high levels of educational attainment themselves. It is possible that accuracy rates would be lower in less elite schools, so conducting a study like this one with a more heterogeneous school sample might be desirable in the future.

Lastly, it is important to bear in mind that the proxy data analyzed in this paper were collected via Internet interviews. The reporting process is likely to vary when respondents are asked questions using oral interview formats, such as the telephone, because the interviewee will not be able to see all of the response categories when answering the question. It is therefore worthwhile to conduct studies involving visual presentation of the question sequence to assess its performance.

BIBLIOGRAPHY

Amato, P. R., & Ochiltree, G. (1987). Interviewing Children about Their Families: A Note on Data Quality. *Journal of Marriage and the Family , 49*, 669-675.

Andresen, E. M., Vahle, V. J., & Lollar, D. (2001). Proxy Reliability: Health-Related Quality of Life (HRQoL) Measures for People with Disability. *Quality of Life Research , 10*, 609-619.

Bassett, S. S., Magaziner, J., & Hebel, J. R. (1990). Reliability of Proxy Response on Mental Health Indices for Aged, Community-Dwelling Women. *Psychology and Aging , 5* (1), 127-132.

Berk, M. L., Horgan, C. M., & Meyers, S. M. (1986). The Reporting of Stigmatizing Health Conditions: A Comparison of Proxy and Self-Reporting. *Journal of Economic and Social Measurement , 14*, 197-205.

Bickart, B. A., Blair, J., Menon, G., & Sudman, S. (1990). Cognitive Aspects of Proxy Reporting of Behavior. *Advances in Consumer Research , 17*, 198-206.

Bokemeier, J., & Monroe, P. (1983, August). Continued Reliance on One Respondent in Family Decision-Making Studies: A Content Analysis. *Journal of Marriage and the Family* , 645-652.

Borus, M. E., & Nestel, G. (1973). Response Bias in Reports of Father's Education and Socioeconomic Status. *Journal of the American Statistical Association , 68* (344), 816-820.

Boyle, C. A., Brann, E. A., & Group, S. C. (1992). Proxy Respondents and the Validity of Occupational and Other Exposure Data. *American Journal of Epidemiology , 136* (6), 712-721.

Briscoe, M. E. (1984). Research Note: Proxy Responses in Health Surveys: A Methodological Issue. *Sociology of Health and Illness , 6* (3), 359-365.

Calsyn, R. J., Morse, G. A., Klinkenberg, W. D., & Trusty, M. L. (1997). Reliability and Validity of Self-Report Data of Homeless Mentally Ill Individuals. *Evaluation and Program Planning , 20* (1), 47-54.

Chang, P.-C., & Yeh, C.-H. (2005). Agreement Between Child Self- Report and Parent Proxy-Report to Evaluate Quality of Life in Children with Cancer. *Psycho-Oncology , 14*, 125-134.

Chen, M.-H., Hsieh, C.-L., Mao, H.-F., & Huang, S.-L. (2007). Differences Between Patient and Proxy Reports in the Assessment of Disability After Stroke. *Clinical Rehabilitation , 21*, 351-356.

Chermack, S. T., Singer, K., & Beresford, T. P. (1998). Screening for Alcoholism Among Medical Inpatients: How Important is Corrobration of Patient Self-Report? *Alcoholism: Clinical and Experimental Research , 22* (7), 1393-1398.

Clarridge, B. R., & Massagli, M. P. (1989). The Use of Female Spouse Proxies in Common Symptom Reporting. *Medical Care , 27* (4), 352-366.

Cohen, R. S., & Orum, A. M. (n.d.). Parent-Child Consesus on Socioeconomic Data Obtained from Sample Surveys. *Public Opinion Quarterly* .

Cole, J. R., & Zuckerman, H. (1984). The Productivity Puzzle: Persistence and Change in Patterns of Publication Among Men and Women Scientists. In M. L. Maehr, & M. W. Steinkamp, *Advances in Motivation and Achievement.* Greenwich, CT: JAI.

Costa, P. T., & McCrae, R. R. (1988). Personality in Adulthood: A Six-Year Longitudinal Study of Self-Reports and Spouse Ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology , 54* (5), 853-863.

Cremeens, J., Eiser, C., & Blades, M. (2006). Factors Influencing Agreement between Child Self-Report and Parent Proxy-Reports on the Pediatric Quality of Life Inventory 4.0 (PedsQL) Generic Core Scales. *Health and Quality of Life Outcomes , 4* (58).

Dorman, P. J., Waddell, F., Slattery, J., Dennis, M., & Sandercock, P. (1997). Are Proxy Assessments of Health Status After Stroke with the EuroQol Questionnaire Feasible, Accurate and Unbiased? *Stroke , 28* (10), 1883-1887.

Douglas, S. P., & Wind, T. (1978). Examining Family Role and Authority Patterns: Two Methodological Issues. *Journal of Marriage and the Family , 40*, 35-47.

Epstein, A., Hall, J. A., Tognetti, J., Son, L. H., & Jr., L. C. (1989). Using Proxies to Evaluate Quality of Life: Can They Provide Valid Information about Patients' Health Status and Satisfaction with Medical Care? *Medical Care , 27* (3), S91-S98.

Gilpin, E. A., Pierce, J. P., Cavin, S. W., Berry, C. C., Evans, N. J., Johnson, M., et al. (1994). Estimates of Population Smoking Prevalence: Self- vs. Proxy Reports of Smoking Status. *American Journal of Public Health , 84* (10), 1576-1579.

Godwin, D. D., & Scanzoni, J. (1989). Couple Consensus During Marital Joint Decision-Making: A Context, Process, Outcome Model. *Journal of Marriage and the Family , 51*, 943-956.

Goldin, C. (1997). Career and Family: College Women Look to the Past. In F. D. Blau, & R. G. Ehrenberg, *Gender and Family Issues in the Workplace.* New York: Russell Sage Foundation.

Graham, P., & Jackson, R. (1993). Primary versus Proxy Respondents: Comparability of Questionnaire Data on Alcohol Consumption. *American Journal of Epidemiology , 138* (6), 443-452.

Hahn, R. A., Truman, B. I., & Barker, N. D. (1996). Identifying Ancestry: The Reliability of Ancestral Identification in the United States by Self, Proxy, Interviewer, and Funeral Director. *Epidemiology , 7* (1), 75-80.

Hatch, M. C., Misra, D., Kabat, G. C., & Kartzmer, S. (1991). Proxy Respondents in Reproductive Research: A Comparison of Self- and Partner-Reported Data. *American Journal of Epidemiology , 133* (8), 826-831.

Hays, R. D., Vickrey, B. G., Hermann, B. P., Perrine, K., Cramer, J., Meador, K., et al. (1995). Agreement between Self Reports and Proxy Reports of Quality of Life in Epilepsy Patients. *Quality of Life Research , 4*, 159-168.

Highton, B. (2005). Self-Reported versus Proxy-Reported Voter Turnout in the Current Population Survey. *Public Opinion Quarterly , 69* (1), 113-123.

Jokovic, A., Locker, D., & Guyatt, G. (2004). How Well Do Parents Know Their Children? Implications for Proxy Reporting of Child Health-Related Quality of Life. *Quality of Life Research , 13*, 1297-1307.

Kain, J. F., & Quigley, J. M. (1972). Note on Owner's Estimate of Housing Value. *Journal of the American Statistical Association , 67* (340), 803-806.

Kayser, B. D., & Summers, G. F. (1973). The Adequacy of Student Reports of Parental SES Characteristics. *Sociological Methods & Research , 1* (3), 303-315.

Klinkenberg, M., Smit, J. H., Deeg, D. J., Willems, D. L., Onwuteaka, B. D., & Wal, G. v. (2003). Proxy Reporting in After-Death Interviews: The Use of Proxy Respondents in Retrospective Assessment of Chronic Diseases and Symptom Burden in the Terminal Phase of Life. *Palliative Medicine , 17*, 191-201.

Lien, N., Friestad, C., & Klepp, K. I. (2001). Adolescents' Proxy Reports of Parents' Socioeconomic Status: How Valid Are They? *Journal of Epidemiology and Community Health , 55*, 731-737.

Long, K., Sudha, S., & Mutran, E. J. (1998). Elder-Proxy Agreement Concerning the Functional Status and Medical History of the Older Person: The Impact of Caregiver Burden and Depressive Symptomatology. *Journal of the American Geriatrics Society , 46* (9).

Looker, E. D. (1989). Accuracy of Proxy Reports of Parental Status Characteristics. *Sociology of Education , 62*, 257-276.

Lyon, J. L., Egger, M. J., Robison, L. M., French, T. K., & Gao, R. (1992). Misclassification of Exposure in a Case-Control Study: The Effects of Different Types of Exposure and Different Proxy Respondents in a Study of Pancreatic Cancer. *Epidemiology , 3* (3), 223-231.

Magaziner, J., Bassett, S. S., Hebel, J. R., & Gruber-Baldini, A. (1996). Use of Proxies to Measure Health and Functional Status in Epidemiologic Studies of Community-Dwelling Women Aged 65 Years or Older. *American Journal of Epidemiology , 143* (3), 283-292.

Magaziner, J., Simonsick, E. M., Kashner, T. M., & Hebel, J. R. (1988). Patient-Proxy Response Comparability on Measures of Patient Health and Functional Status. *Journal of Clinical Epidemiology , 41* (11), 1065-1074.

Magaziner, J., Zimmerman, S. I., Gruber-Baldini, A. L., Hebel, J. R., & Fox, K. M. (1997). Proxy Reporting in Five Areas of Functional Status: Comparison with Self-Reports and Observations of Performance. *American Journal of Epidemiology , 146* (5), 418-428.

Maisto, S. A., Sobell, M. B., & Sobell, L. C. (1982). Reliabiilty of Self-Reports of Low Ethanol Consumption by Problem Drinkers over 18 Months of Follow-Up. *Drug and Alcohol Dependence , 9*, 273-278.

Mare, R. D., & Mason, W. M. (1980). Children's Reports of Parental Socioeconomic Status: A Multiple Group Measurement Model. *Sociological Methods & Research , 9* (2), 178-198.

Martin, J., & Butcher, B. (1982). The Quality of Proxy Information--Some Results from a Large-Scale Study. *The Statistician , 31* (4), 293-319.

Massagli, M. P., & Hauser, R. M. (1983). Response Variability in Self- and Proxy Reports of Paternal and Filial Socioeconomic Characteristics. *AJS , 89* (2), 420-431.

Medin, K., & Wilson, B. (1988). Trends in Agricultural Statistics--An Outline of Development Work at Statistics Sweden. *Journal of Official Statistics , 4* (2), 173-190.

Menon, G., Bickart, B., Sudman, S., & Blair, J. (1995). How Well Do You Know Your Partner? Strategies for Formulating Proxy-Reports and Their Effects on Convergence to Self-Reports. *Journal of Marketing Research , 32*, 75-84.

Miller, R. E., Massagli, M. P., & Clarridge, B. R. *Quality of Proxy vs. Self-Reports: Evidence from a Health Survey with Repeated Measures.*

Moinpour, C. M., Lyons, B., Schmidt, S. P., Chansky, K., & Patchell, R. A. (2000). Substituting Proxy Ratings for Patient Ratings in Cancer Clinical Trials: An Analysis Based on a Southwest Oncology Group Trial in Patients with Brain Metastases. *Quality of Life Research , 9*, 219-231.

Mosely, R. R., & Wolinsky, F. D. (1986). The Use of Proxies in Health Surveys: Substative and Policy Implications. *Medical Care , 24* (6), 496-510.

Moskowitz, D. S. (1986). Comparison of Self-Reports, Reports by Knowledgeable Informants, and Behavioral Observation Data. *Journal of Personality , 54* (1), 294-317.

Navarro, A. M. (1999). Smoking Status by Proxy and Self Report: Rate of Agreement in Different Ethnic Groups. *Tobacco Control , 8*, 182-185.

Nekolaichuk, C. L., Bruera, E., Spachynski, K., MacEachern, T., Hanson, J., & O'Maguire, T. (19999). A Comparison of Patient and Proxy Symptom Assessments in Advanced Cancer Patients. *Palliative Medicine , 13*, 311-323.

Nelson, L. M., Jr., W. T., Koepsell, T. D., & Belle, G. v. (1990). Proxy Repondents in Epidemiologic Research. *Epidemiologic Research , 12*, 71-85.

Nelson, L. M., Longstreth, W. T., Koepsell, T. D., Checkoway, H., & van Belle, G. (1994). Completeness and Accuracy of Interview Data from Proxy Respondents: Demographic, Medical, and Life-Style Factors. *Epidemiology , 5* (2), 204-217.

Novella, J. L., Jochum, C., Jolly, D., Morrone, I., Ankri, J., Bureau, F., et al. (2001). Agreement Between Patients' and Proxies' Reports of Quality of Life in Alzheimer's Disease. *Quality of Life Research , 10*, 443-452.

Olson, D. H. (1969). The Measurement of Family Power by Self-Report and Behavioral Methods. *Journal of Marriage and the Family* , 545-550.

Pierre, U., Wood-Dauphinee, S., Korner-Bitensky, N., Gayton, D., & Hanley, J. (1998). Proxy Use of the Canadian SF-36 in Rating Health Status of the Disabled Elderly. *Journal of Clinical Epidemiology , 51* (11), 983-990.

Pyles, M. K., Stolz, H. R., & Macfarlane, J. W. (1935). The Accuracy of Mothers' Reports on Birth and Development Data. *Child Development , 6* (2), 165-176.

Rasinski, K. A. (1989). *How Seriously Should We Take Children's Judgments About the Seriousness of Problems in Their Schools? A Validation of Children's Survey Responses.* NORC.

Ronen, G. M., Streiner, D. L., Rosenbaum, P., & Network, C. P. (2003). Health-Related Quality of Life in Children with Epilepsy: Development and Validation of Self-Report and Parent Proxy Measures. *Epilepsia , 44* (4), 598-612.

Rothman, M. L., Hedrick, S. C., Bulcroft, K. A., Hickam, D. H., & Rubenstein, L. Z. (1991). The Validity of Proxy-Generated Scores as Measures of Patient Health Status. *Medical Care , 29* (2), 115-124.

Russell, J. A. (1989). Measures of Emotion. In R. Plutchik, & H. Kellerman (Eds.), *Emotion: Theory, Research, and Experience* (pp. 83-111). San Diego: Academic Press, Inc.

Schwarz, N., & Wellens, T. (1997). Cognitive Dynamics of Proxy Responding: The Diverging Perspectives of Actors and Observers. *Journal of Official Statistics , 13* (2), 159-179.

Shaw, C., McColl, E., & Bond, S. (2001). Functional Abilities and Continence: The Use of Proxy Respondents in Research Involving Older People. *Quality of Life Research , 9*, 1117-1126.

Shields, M. Proxy Reporting in the National Population Health Survey. *Health Reports , 12* (1), 21-39.

Siemiatycki, J., Campbell, S., Richardson, L., & Aubert, D. (1984). Quality of Response in Different Population Groups in Mail and Telephone Surveys. *American Journal of Epidemiology , 120* (2), 302-313.

Sneeuw, K. C., Aaronson, N. K., Haan, R. J., & Limburg, M. (1997). Assessing Quality of Life After Stroke: The Value and Limitations of Proxy Ratings. *Stroke , 28*, 1541-1549.

Sneeuw, K. C., Aaronson, N. K., Osoba, D., Muller, M. J., Hsu, M.-A., Yung, W. K., et al. (1997). The Use of Significant Others as Proxy Raters of the Quality of Life of Patients with Brain Cancer. *Medical Care , 35* (5), 490-506.

Sneeuw, K. C., Aaronson, N. K., Sprangers, M. A., Detmar, S. B., Wever, L. D., & Schornagel, J. H. (1998). Comparison of Patient and Proxy EORTC QLQ-C30 Ratings in Assessing the Quality of Life of Cancer Patients. *Journal of Clinical Epidemiology , 51* (7), 617-631.

Sobell, L. C., Toneatto, T., Sobell, M. B., Leo, G. I., & Johnson, L. (1992). Alcohol Abusers' Perceptions of the Accuracy of Their Self-Reports of Drinking: Implications for Treatment. *Addictive Behaviors , 17*, 507-511.

Sudman, S., Bickart, B., Blair, J., & Menon, G. The Effect of Participation Level on Reports of Behavior and Attitudes by Proxy Reporters.

Tedin, K. L. (1976). On the Reliability of Reported Political Attitudes. *American Journal of Political Science , 20* (1), 117-124.

Theunissen, N. C., Vogels, T. G., Koopman, H. M., Verrips, G. H., Zwinderman, K. A., Verloove-Vanhorick, S. P., et al. (1998). The Proxy Problem: Child Report Versus Parent Report in Health-Related Quality of Life Research. *Quality of Life Research , 7*, 387-397.

Todorov, A., & Kirchner, C. (2000). Bias in Proxies' Reports of Disability: Data from the National Health Interview Survey on Disability. *American Journal of Public Health , 90* (8), 1248-1253.

Walker, A. M., Velema, J. P., & Robins, J. M. (1988). Analysis of Case-Control Data Derived in Part from Proxy Respondents. *American Journal of Epidemiology , 127* (5), 905-914.

Williams, R., & Thomson, E. (1985). Can Spouses be Trusted? A Look at Husband/Wife Proxy Reports. *Demography , 22* (1), 115-123.

Wilson, K. A., Dowling, A. J., Abdolell, M., & Tannock, I. F. (2000). Perception of Quality of Life by Patients, Partners and Treating Physicians. *Quality of Life Research , 9*, 1041-1052.

Yeh, C.-H., Chang, C.-W., & Chang, P.-C. (2005). Evaluating Quality of Life in Children with Cancer Using Children's Self-Reports and Parent-Proxy Reports. *Nursing Research , 54* (5), 354-362.

Table 1. Necessary Features of a Proxy Accuracy Study in Previous Research.

Last Name of Authors	Year	Both Targets/Proxies Interviewed	Representative Samples of the Same Populations	Identical Questions	External Measure of Accuracy
Amato & Ochiltree	1987			x	x
Andersen et al.	1973				
Andresen, Vahle & Lollar	2001		x	x	x
Balamuth	1965				
Bassett, Magaziner & Hebel	1990				x
Berk, Horgan & Meyers	1986		x	x	x
Bielby & Hauser	1977		x		x
Borus & Nestel	1973		x		x
Boyle, Brann & Group	1992		x	x	x
Briscoe	1984			x	x
Calsyn et al.	1997		x	x	x
Cannel & Fowler	1963				
Chang & Yeh	2005				x
Chen et al.	2007				x
Chermack, Singer & Beresford	1998				x
Clarridge & Massagli	1989				x
Cobb et al.	1956				
Cohen & Orum	1972				x
Costa & McCrae	1988		x	x	x
Cremeens, Eiser & Blades	2006			x	x
Dorman et al.	1997			x	x
Douglas & Wind	1978				x
Epstein et al.	1989				x
Gilpin et al.	1994		x	x	x
Godwin & Scanzoni	1989				x
Graham & Jackson	1993			x	x
Hahn, Truman & Barker	1996				x
Hatch et al.	1991				x
Hauser et al.	1983				x
Hays et al.	1995				x
Herrmann	1985				x
Highton	2005	x	x		x
Jokovic, Locker & Guyatt	2004			x	x
Kain & Quigley	1972		x		x
Kayser & Summers	1973				x
Kerckhoff et al.	1973				x
Klinkenberg et al.	2003	x	x		
Lien, Friestad & Klepp	2001			x	x
Long, Sudha & Mutran	1998				x
Lyon et al.	1992				x
Magaziner et al.	1997				
Magaziner et al.	1988	x	x	x	x
Magaziner et al.	1996				x
Maisto, Sobell & Sobell	1982			x	x
Mare & Mason	1980				x

Massagli & Hauser	1983				x
Menon et al.	1995				x
Miller, Massagli & Clarridge			x	x	x
Moinpour et al.	2000				x
Mosely II & Wolinsky	1986	x	x		x
Navarro	1999		x	x	x
Nekolaichuk et al.	1999				x
Nelson et al.	1994	x	x		x
Niemi	1974				x
Novella et al.	2001				x
Olson	1969				x
Pickle & Brown	1983				x
Pierre et al.	1998				x
Pyles, Stolz & Macfarlane	1935	x			x
Rasinski	1989	x	x	x	
Ronen et al.	2003		x	x	x
Rothman et al.	1991				x
Schwarz & Wellens	1997				x
Shaw, McColl & Bond	2001				x
Shields	1982	x	x		x
Siemiatycki et al.	1984		x		
Sneeuw et al.	1997	x	x		x
Sneeuw et al.	1997				x
Sneeuw et al.	1998				x
Sudman et al.					x
Tedin	1976				x
Theunissen et al.	1998				x
Thomspon & Tauber	1957				
Todorov & Kirchner	2000	x	x		x
Walker, Valema & Robins	1988	x	x		x
Williams & Thomson	1985			x	x
Wilson et al.	2000				x
Yeh, Chang & Chang	2005				x

An “x” in a particular column indicates that the study lacked the feature in question. The gray rows represent studies that fulfilled all requirements.

Table 2. Univariate Statistics of Respondents and Field(s) of Study by Group (Stanford/UVA Alumni Telephone Interviews).

	Stanford		University of Virginia	
	7 Categories	**10 Categories**	**7 Categories**	**10 Categories**
N of Respondents				
Number	328	306	266	307
Response Rate	11.5%	[a]	10.8%	[a]
Reverse Category Order	48.2%	46.1%	50.4%	45.0%
7 Categories[b]				
Biological, agricultural, physical, or related sciences	23.2%		9.0%	
Health, nursing, or medical fields	1.5		7.9	
Engineering, computer, mathematical, or related sciences	19.2		15.8	
Psychology, economics, or other social sciences, except history	31.4		22.2	
History, arts, or humanities	21.7		25.6	
Business, communication or education	3.1		18.1	
Some other field(s)	0.0		1.5	
TOTAL	100.0		100.0	
10 Categories[b]				
Biological, agricultural, or related sciences		19.0%		6.2%
Health, nursing, or medical fields		1.6		5.9
Computer, mathematical, or related sciences		4.6		3.0
Engineering		12.8		17.7
Physical or related sciences, including earth sciences		3.3		2.3
Psychology, economics, or other social sciences, except history		30.1		22.3
Business or communication		2.6		13.4
Education		1.0		2.6
History, arts, or humanities		24.5		24.6
Some other field(s)		0.7		2.0
TOTAL		100.0		100.0
Average Number of Classifications	1.30	1.29	1.28	1.29
(Standard Deviation)	(0.62)	(0.72)	(0.59)	(0.60)
Degree Year	1979.4	1977.6	1986.2	1986.0
(Standard Deviation)	(14.94)	(13.94)	(11.59)	(11.29)
Female	46.7	45.1	53.0	45.6

Notes: [a] Response rate for total survey instead of by condition. [b] Classified according to data reported by the Stanford University Registrar's Office.

Table 3. Open Question Reporting Errors by School and Response Condition (Stanford/UVA Alumni Telephone Interviews).

	Stanford				University of Virginia			
	7 Original	**10 Original**	**7 Reverse**	**10 Reverse**	**7 Original**	**10 Original**	**7 Reverse**	**10 Reverse**
Correct Report	90.6%	93.3%	95.6%	92.2%	82.6%	85.6%	82.1%	82.6%
Misreport	4.1	3.0	0.6	1.4	4.6	4.2	3.0	4.4
Over-report	5.3	3.0	3.8	6.4	5.3	6.0	7.5	7.3
Under-report	0	0.6	0	0	7.6	4.2	7.5	5.8
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Correctly Report for Science and Engineering	99.4%	97.6%	99.4%	97.2%	96.2%	96.4%	94.8%	95.7%
Misreport into Science and Engineering	0.6	1.8	0.6	2.1	0.8	2.4	1.5	3.6
Misreport out of Science and Engineering	0	0.6	0	0.7	3.0	1.2	3.7	0.7
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	170	165	158	141	132	167	134	138

Table 4. Closed Question Classification Errors by School and Response Condition (Stanford/UVA Alumni Telephone Interviews).

	Stanford				University of Virginia			
	7 Original	**10 Original**	**7 Reverse**	**10 Reverse**	**7 Original**	**10 Original**	**7 Reverse**	**10 Reverse**
Correct Classification	71.8%	70.9%	66.5%	68.1%	69.7%	64.1%	67.9%	58.7%
Misclassification	9.4	10.9	13.3	12.8	15.9	19.2	17.2	18.1
Correct Classification plus Erroneous Additional Classifications	14.7	15.2	15.2	14.2	14.4	16.8	14.9	23.2
Incomplete Correct Classification	4.1	3.0	5.1	5.0	0	0	0	0
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Correctly Classified for Science and Engineering	96.5%	93.3%	87.3%	89.4%	88.6%	84.4%	82.8%	87.0%
Misclassified into Science and Engineering	1.8	2.4	3.8	2.1	3.0	5.4	5.2	4.4
Misclassified out of Science and Engineering	1.8	4.2	8.9	8.5	8.3	10.2	11.9	8.7
TOTAL	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
N	170	165	158	141	132	167	134	138

Table 5. Logistic Regression Predicting Closed Response Classification Errors (Stanford/UVA Alumni Telephone Interviews).

	Stanford			University of Virginia		
	All Errors	**Misclassification Errors**	**SE Error**	**All Errors**	**Misclassification Errors**	**SE Error**
Open Question Error						
No Open Error [a]						
Misreport	1.47**	.44	.29	1.69***	1.95***	1.15*
Over-report	1.51***	1.58***	2.05***	1.07**	.004	.66
Under-report	[b]	[b]	[b]	1.28**	2.18***	.76[†]
Class Year	.01[†]	.003	-.004	.01[†]	.003	.001
Female	.34[†]	.50[†]	.07	-.09	.19	.10
Response Condition						
7-Original [a]						
10-Original	.16	.29	.81	.31	.38	.40
7-Reverse	.38	.49	1.52**	.10	.16	.50
10-Reverse	.26	.36	1.20*	.54*	.26	.17
Intercept	-24.65***	-2.84	3.99	-29.06***	-9.29	-4.12
X^2 Statistic	28.87***	17.14*	29.56***	41.59***	50.43***	11.27
DF	7	7	7	8	8	8
N	633	633	633	571	571	571

[†].1 *.05 **.01 ***.001

[a] Reference Category [b] predicts perfect failure

SE Error is Science and Engineering Reporting Error.

Table 6. Univariate Statistics of Respondents and Field(s) of Study by Group (Stanford Alumni-Parents Matching).

	Alumni		Parents	
	7 Categories	**10 Categories**	**7 Categories**	**10 Categories**
N of Respondents				
Number	1,000	1,013	832	900
Response Rate	51.0%	49.1%	46.6%	48.1%
Reported on Two Children	--	--	9.4%	9.3%
Alumnus-Parent Matching	491	499	491	499
	49.1%	49.3%	59.0%	55.4%
7 Categories[a]				
Biological, agricultural, physical, or related sciences	20.3%		19.9%	
Health, nursing, or medical fields	0.0		0.0	
Engineering, computer, mathematical, or related sciences	20.1		20.9	
Psychology, economics, or other social sciences, except history	40.2		40.7	
History, arts, or humanities	17.0		16.5	
Business, communication or education	2.0		1.7	
Some other field(s)	0.4		0.3	
TOTAL	100.0		100.0	
10 Categories[a]				
Biological, agricultural, or related sciences		16.4%		15.8%
Health, nursing, or medical fields		0.0		0.1
Computer, mathematical, or related sciences		4.3		3.8
Engineering		15.3		16.1
Physical or related sciences, including earth sciences		4.6		3.8
Psychology, economics, or other social sciences, except history		41.2		40.9
Business or communication		1.4		1.7
Education		0.0		0.0
History, arts, or humanities		16.5		17.2
Some other field(s)		0.3		0.6
TOTAL		100.0		100.0
Average Number of Classifications	1.36	1.52	1.26	1.30
(Standard Deviation)	(0.68)	(0.87)	(0.61)	(0.64)
Alumni Degree Year	1997.1	1997.5	1997.3	1997.7
(Standard Deviation)	(8.78)	(8.52)	(8.25)	(8.11)
Female	53.9	55.9	45.3	45.4
Parent Advanced Degree			24.7	25.0

Note: [a] Classified according to data reported by the Stanford University Registrar's Office.

Table 7. Open Question Reporting Errors by School and Response Condition (Stanford Alumni-Parents Matching).

	Alumni		Parents	
	7 Categories	**10 Categories**	**7 Categories**	**10 Categories**
Correct Report	96.8%	96.6%	89.9%	87.7%
Misreport	0.6	0.3	5.5	7.6
Over-report	2.2	2.4	3.9	3.3
Under-report	0.4	0.7	0.8	1.5
TOTAL	100.0%	100.0%	100.0%	100.0%
Correctly Report for Science and Engineering	99.3%	99.5%	98.0%	97.2%
Misreport into Science and Engineering	0.3	0.3	1.4	2.1
Misreport out of Science and Engineering	0.4	0.2	0.6	0.7
TOTAL	100.0%	100.0%	100.0%	100.0%
N	989	996	880	947

Table 8. Closed Question Classification Errors by School and Response Condition (Stanford Alumni-Parent Matching).

	Alumni		Parents	
	7 Categories	**10 Categories**	**7 Categories**	**10 Categories**
Correct Classification	69.8%	58.3%	64.8%	56.2%
Misclassification	11.5	14.4	18.2	24.9
Correct Classifications plus Erroneous Additional Classifications	14.1	22.1	9.9	12.6
Incomplete Correct Classifications	4.6	5.1	7.0	7.3
TOTAL	100.0%	100.0%	100.0%	100.0%
Correctly Classified for Science and Engineering	93.2%	93.4%	88.7%	87.8%
Misclassified into Science and Engineering	5.6	5.6	10.0	10.7
Misclassified out of Science and Engineering	1.2	1.0	1.3	1.5
TOTAL	100.0%	100.0%	100.0%	100.0%
N	1000	1013	910	983

Table 9. Logistic Regression Predicting Closed Response Classification Errors (Stanford Alumni-Parents Matching).

	Alumni			Parents		
	All Errors	**Misclassification Errors**	**SE Errors**	**All Errors**	**Misclassification Errors**	**SE Errors**
Open Question Error						
No Open Error [a]						
Misreport	1.57*	-.08	[b]	2.00***	1.61***	1.78***
Over-report	1.53***	.53	1.08**	1.35***	-.12	0.38
Under-report	1.11[†]	1.72**	1.78[†]	.89*	1.35**	1.04*
Class Year	.03***	.02*	.01	.01	.0004	-.02*
Female (Alumnus)	.12	.19	.63**	.29**	.07	.48***
Response Condition						
7-Original [a]						
10-Original	.50***	.24[†]	-.06	.34**	.30*	.01
Female (Parent)				.001	.05	.26[†]
Multiple Child Response				-.01	-.05	-.31
Advanced Degree (Parent)				-.11	-.25[†]	-.42*
Intercept	-64.38***	-41.22*	-20.00	-19.45	-2.50	39.43
X^2 Statistic	91.24***	20.29**	23.06***	141.25***	85.83***	100.18***
DF	6	6	5	9	9	9
N	2013	2013	2004	1892	1892	1892

[†].1 *.05 **.01 ***.001

[a] Reference Category [b] predicts perfect failure

SE Error is Science and Engineering Reporting Error.

Table 10. Parent Errors by Whether Alumni Responded (Stanford Alumni-Parent Matching).

	Closed		Open	
	Response	**Non-Response**	**Response**	**Non-Response**
Parents				
Correct Report	60.7%	59.9%	89.3%	88.1%
Misreport	20.4	22.0	5.4	7.9
Over-report	11.6	11.0	3.8	3.4
Under-report	7.3	7.1	1.6	0.7
Total Percent	100%	100%	100%	100.0%
Total N	990	903	961	866
X^2 (3)	0.85		7.39	

Table 11. Percent of Parents Open Question Report Errors by Alumni Classification Errors (Stanford Alumni-Parent Matching).

	Alumni			
	Correct Report	Misreport	Over-report	Under-report
Parents				
Correct Report	90.1%	100.0%	68.4%	33.3%
Misreport	5.0	0	21.1	16.7
Over-report	3.6	0	10.5	0
Under-report	1.3	0	0	50.0
TOTAL N	921	4	19	6
TOTAL %	100%	100%	100%	100%

Table 12. Percent of Parent Closed-Question Classifications Errors by Alumni Classification Errors (Stanford Alumni-Parent Matching).

	Alumni			
	Correct Classification	Misclassification	Correct Classification + Erroneous Classification	Incomplete Correct Classifications
Parents				
Correct Classification	78.9%	19.1%	44.6%	2.0%
Misclassification	9.5	58.8	29.0	24.5
Correct Classification + Erroneous Classifications	10.3	8.4	20.4	4.1
Incomplete Correct Classifications	1.4	13.7	5.9	69.4
TOTAL N	624	131	186	49
TOTAL %	100.0	100.0	100.0	100.0

Table 13. Agreement by Alumni and Parents (Stanford Alumni-Parent Matching).

	All Parents		Parents Making Any Error		Parents Making Misclassification Error	
	% Agree	Cohen's K	% Agree	Cohen's K	% Agree	Cohen's K
7 Categories						
Biological, agricultural, physical, or related sciences	94.9	.84***	88.0	.74***	86.4	.57***
Health, nursing, or medical fields	92.0	.22***	81.7	.20**	89.8	.52***
Engineering, computer, mathematical, or related sciences	96.5	.92***	91.5	.77***	93.2	.81***
Psychology, economics, or other social sciences, except history	85.5	.68***	65.1	.27***	54.6	.05
History, arts, or humanities	86.9	.66***	72.6	.28***	63.6	.17*
Business, communication or education	95.9	.59***	92.0	.46***	95.6	.58***
Some other field(s)	87.1	.24***	69.7	.16*	60.2	.14†
10 Categories						
Biological, agricultural, or related sciences	92.8	.76***	87.8	.69***	90.4	.62***
Health, nursing, or medical fields	94.4	.31***	89.2	.32***	93.0	.56***
Computer, mathematical, or related sciences	92.8	.73***	89.2	.73***	93.9	.86***
Engineering	94.8	.83***	88.7	.57***	83.3	.24***
Physical or related sciences, including earth sciences	94.4	.59***	92.0	.28***	93.9	-.03
Psychology, economics, or other social sciences, except history	82.8	.62***	68.1	.33***	61.4	.17**
Business or communication	96.4	.51***	96.7	.57***	99.1	--
Education	96.6	.09*	93.4	.10†	95.6	.27***
History, arts, or humanities	84.8	.63***	75.1	.41***	66.7	.29***
Some other field(s)	89.2	.21***	79.8	.21***	71.1	.16*

Table 14. Logistic Regression Predicting Closed Response Classification Errors by Parents Matched to Alumni Errors (Stanford Alumni-Parent Matching).

	All Errors	Parents Misclassification Errors	SE Error
Alumni Closed Error			
No Error [a]			
Misclassification	2.88***	2.70***	1.94***
Over-classification	1.64***	1.43***	1.01***
Under-classification	5.24***	1.04***	-1.49
Open Question Error			
No Open Error [a]			
Misreport	1.91***	1.63***	1.61***
Over-report	1.92***	-.04	-.04
Under-report	.92	1.66**	.94
Class Year	-.01	-.01	-.04**
Female (Alumnus)	.11	-.13	.46*
Response Condition			
7-Original [a]			
10-Original	.09	.05	-.24
Female (Parent)	.08	-.03	.34
Multiple Child Response	-.06	-.10	-.36
Advanced Degree (Parent)	-.10	-.42[†]	-.64*
Intercept	9.47	14.00	71.39
X^2 Statistic	344.88***	191.63***	100.61***
DF	12	12	12
N	990	990	990

[†].1 *.05 **.01 ***.001

[a] Reference Category [b] predicts perfect failure

SE Error is Science and Engineering Reporting Error.