

SURVEY NONRESPONSE ISSUES

Addressing Survey Nonresponse Issues: Implications for ATE Principal Investigators,
Evaluators, and Researchers

Wayne W. Welch, University of Minnesota
Ashly N. Barlau, University of Colorado, Boulder

Author Note

Wayne W. Welch, Professor Emeritus, University of Minnesota and Consultant, Rainbow Research, Inc., Minneapolis, MN

Ashly N. Barlau, University of Colorado, Boulder

This material is based upon work supported by the National Science Foundation under Grants No. 0832874 and No. 1132099

Correspondence concerning this article should be addressed to Wayne Welch, 621 W. Lake St., Minneapolis, MN 55408

Contact: wwelch@umn.edu

Abstract

Many ATE principal investigators (PIs), evaluators, and researchers, use mailed or emailed surveys to help them assess the effectiveness of their work. However, a common problem is low response rate and the potential that it creates for nonresponse bias. This is the bias that occurs when those who do not respond to a survey are different in some systematic way, from those who do answer the survey. If this occurs, serious questions may be raised about the validity of the findings.

The purpose of this report is to raise awareness and present ways to address nonresponse problems among those that use surveys in their studies of ATE grants. This includes PIs, evaluators, researchers, and those who may be evaluating the full program. We summarize the research on nonresponse issues, present generally acceptable standards for response rates, offer suggestions on how to increase response rates, describe ways to check for nonresponse bias, and apply these methods to a research study of the Advanced Technological Education (ATE) program sponsored by the National Science Foundation (NSF).

We used a mailed survey to gather information about the impact and sustainability of the ATE program. Although we had a high response rate, we decided to check for nonresponse bias. We used three methods to investigate the problem; we compared responders with the total population and compared responders with nonresponders on five background characteristics. We also compared early responders with late responders on two scales, an ATE Impact Scale and an ATE Sustainability Scale. We used late responders as a surrogate for nonresponders.

We found a slightly higher response rate for the larger center grants when compared to projects. However, we found no differences in the actual survey responses between early responders and late responders, our proxy for nonresponders. This led us to believe we did not have a nonresponse bias in our results.

We believe that our experience will be useful for those doing research on and evaluation of the ATE program. We hope the suggestions we offer will help improve the validity of their findings.

Addressing Survey Nonresponse Issues: Implications for ATE Principal Investigators, Evaluators, and Researchers

Carrying out research or evaluation studies on programs such as the Advanced Technological Education (ATE) program supported by the National Science Foundation (NSF) often involves the use of questionnaires or surveys to obtain information from project grantees. Surveys may be sent by surface mail but in recent years, email and web-based surveys have become quite popular and many companies are available for hire to do this. Each method has its strengths and limitations but a major concern for all methods is the potential bias caused when subjects do not return the survey. If the nonrespondents' replies had been different from those that did respond, the external validity of the findings might be flawed. This threat is called nonresponse bias. PIs, evaluators, and researchers can address this issue by obtaining high response rates and by identifying the nature of the bias.

In this report, we examine nonresponse issues and describe ways to address these problems. We focus on research and evaluation carried out as part of the ATE program. We summarize the research on nonresponse issues, present generally acceptable standards for response rates, offer suggestions on how to increase response rates, and describe ways to identify nonresponse bias.

We conducted a preliminary review of the literature to determine how investigators handle survey nonresponse. Despite a large number of published survey studies, few addressed the nonresponse issue. Those who have addressed this problem have done so in several ways. One way is to ignore it and many researchers tend to choose this option. However, there are methods that researchers can use (Miller & Smith, 1983; Lindner, Murphy, & Briers, 2001). They can:

(1) Compare respondents to the population. This requires that the researchers have information about the background characteristics of the population, for example, size of the grant or type of institution that received a grant. If there are differences, the results only can be generalized to the respondents. This approach is not very sensitive if there is a high response rate because of the large overlap between the respondents and the population.

(2) Compare respondents to nonrespondents. This method also requires the researchers have background information about both groups. This could be the size of the grant or, perhaps, whether the grant was made to a two-year or four-year institution.

(3) Compare early to late responders. Some research has shown that late responders are similar to nonrespondents so late responders can be used as a proxy for nonrespondents. Actual survey responses are compared to determine if there are differences between the two groups.

(4) Follow-up on a random sample (10% - 20%) of nonrespondents. Usually telephone or personal interviews are used. Questions from the survey are used to guide the interviews.

We used the first three methods for our research on the impact and sustainability of the ATE program. We surveyed the principal investigators (PIs) of active ATE projects and those that had ended during the three-year period prior to our study. Two-hundred sixty-one grantees

met our selection criteria. We mailed the PIs a four-page survey that included questions about sustainability; that is, the continuation of work after NSF funding ended. After three follow-up contacts, we received responses from 212 sites for a response rate of 81.2%. Additional information on the study is available in Welch (2011). Although a response rate of 81% is quite good, there may be differences between the sites that answered the survey and those that did not. We decided to conduct a nonresponse study to determine if there was a nonresponse bias that might threaten the validity of our findings.

Review of the Literature

Questionnaire Response Rates

One way to avoid nonresponse bias is to have high response rates. We examined the literature to determine the level of response obtained by other researchers. We found it to be quite low. In one study of five journals published in the managerial and behavioral sciences, an average of 55% of subjects returned mailed surveys (Baruch, 1999). The response rate for top managers or organizational representatives was only 38%. In addition, over the 20 years covered by the article, the overall response rate dropped from 64% to 48%.

Many researchers have turned to the internet and administer online surveys, but the response rates for these methods are even lower and have decreased over time. This is due the inundation of email messages in general and email surveys in particular (Sheehan, 2001; Kaplowitz, Hadlock, & Levine, 2004).

In another example, a survey mailed to over 3,000 members of the International Technology Education Association (Flowers, 2001) yielded a response rate of 29% with a justification that this response rate was nearly double that of a survey of engineer educators.¹ No other comment about nonresponse was made. This means the researchers had no idea about the opinions of the more than 2,000 people (71%) who failed to return the survey. This creates a high potential for nonresponse bias.

A search of the Educational Resource Information Center (ERIC) was conducted using the identifiers; technology education, response rates, two-year colleges, and both “research reports” and “numerical data.” Twenty-eight articles were identified and these were scanned for reported response rates. Only six studies reported the percentage of replies and they ranged from 15% to 59%. Nearly all of them were follow-up studies of graduates. None of them addressed the nonresponse error potential.

We looked for articles that addressed mailed questionnaires and response rates in the *American Journal of Evaluation* (AJE) and its predecessor, *Evaluation Practice*. Thirty-one articles were found that contained these words and 17 of them reported response rates ranging from 26% to 92%. The median return rate was 59%. Only two of the studies addressed the nonresponse error potential.

¹ Comparing a poor response rate with a very poor one does not make the poor one better!

A few studies compared the response rates between mailed and emailed questionnaires or some combination of these. For example, contrary to their expectations, (Kiernan, Kiernan, Oylar, & Giles, 2005) found that 70% of their subjects returned a web-based survey compared to a 61% return rate for the mailed version of the survey. However, more recent studies have shown lower response rates for web-based or email surveys (Converse, Wolfe, Huang, & Oswald, 2008).

We searched the *Community College Journal of Research and Practice* for articles that contained the word survey or questionnaire. We found 171 in the issues going back to 1997. When we added the search term “nonresponse,” the number dropped to 32. Unfortunately, the library we used at the University of Minnesota does not have online access to that journal so we could not search the articles for more details.

We also examined issues of the *Journal of Science Education and Technology* and identified six articles using the search terms “mailed questionnaire” and “response rates.” One of them (Odom, Settlage, & Pederson, 2002) sent email surveys to 893 members of the Association for the Education of Teachers in Science. Three hundred and seven addresses were returned as undeliverable. A second email request was sent to those who had valid email addresses and 276 members responded. That is 31% of the total membership or 47% of those with valid email addresses. Nothing was said about possible nonresponse bias.

Three other studies in that journal used questionnaires and mentioned response rates. In one, 39 surveys were mailed in a study of the effectiveness of a summer science institute, but only 20 replies (51%) were received. They ignored the 19 that were returned for incomplete addresses and reported a response rate of 80%, 16 out of 20. They do not mention nonresponse error.

Another study examined the workplace skills of graduates of a New Zealand university. They reported that 46% of employers returned the survey and the only justification they give that it was acceptable was that a national commission survey on writing in 2004 only received a 17% response rate!² Finally, we found an article reporting on a post training survey sent to supervisors of graduates of an online training program for law enforcement personnel. Only 27% returned the survey. Again, no mention was made of nonresponse bias.

On the other hand, the Evaluation Center at Western Michigan has been using web-based surveys to gather data about the ATE program for several years. With numerous follow-ups and pressure from program officers at NSF, they have consistently obtained over 90% responses rates (Coryn, Ritchie, & Gullickson, 2006). Their target population is projects and centers that are at least one year old. The sample sizes for the annual survey vary from 76 to 178 for the three years that were examined (2002, 2004, and 2008). The response rates ranged from 92% to 100%.

² See footnote 1.

The center also did a follow-up study of expired projects in 2005 using an online survey with frequent email and telephone follow-ups. The population was 204 sites and during two months of data gathering, 136 responses were received for a 67% response rate.

One can see that response rates are often quite low and little is done about nonresponse bias. We think ATE researchers and evaluators should pay careful attention to obtaining adequate response rates and doing nonresponse bias studies. We will explain how this can be accomplished later in this report.

Addressing Nonresponse Problems

Many texts present suggestions for enhancing response rates. These include such things as a carefully written cover letter on official letterhead that explains the purpose of the survey, avoiding surveys that have a heavy response burden,³ creating visually appealing questionnaires, using simple and easily understood statements, include stamped return-addressed envelopes, including information about a contact person if there are questions, and following-up to nonresponders with letters or emails (Borg & Gall, 1983). We did all of those things and, in addition, offered a small gratuity for their assistance.⁴

There is a substantial amount of literature on nonresponse issues in public opinion surveys (Dillman, Smyth, & Christian 2009), market research (Manfreda, 2008), epidemiological studies (Hartge, 1999), and by the U.S. Census Bureau (Groves & Couper, 1998). These surveys rely on probability sampling to make inferences to large unobserved populations. If the researcher is unable to make contact with the selected households, or if the contacts refuse to participate, incorrect inferences may result. Research in these fields has focused on ways to increase response rates, measuring nonresponse bias, and compensating for nonresponse by imputation and/or weighting (Singer, 2006).

However, the literature on nonresponse issues in educational fields is less extensive. For example, a search of the Educational Resource Information Center (ERIC) using “mailed (survey or questionnaire)” yielded 3,220 publications. When we added the search words “nonresponse” or “nonrespondents,” the number of hits dropped to 115. Adding the term “bias” yielded just 10 articles.

We searched issues of two research journals in science education; the Journal of Research in Science Education (JRST) and Science Education (SCI EDUC) to determine how researchers in that discipline have handled this problem. We used the terms “mailed (survey or questionnaire)” to search articles in SCI EDUC. The search was limited to articles that appeared between the 1963 and 2011, the same years we used to examine JRST articles. We found 459 articles that met our criteria. When we added the terms “nonresponse” and “nonrespondents,” the number of hits dropped to 14 articles.

³ We think survey response time should not be more than 20 minutes.

⁴ We discovered it was difficult to include the \$5 gratuity. It creates many problems for grant administrators because of federal grant restrictions. We think some other kind of token might be better, for example, an ATE embossed pencil or a gift certificate.

We reviewed these articles to discover if they reported response rates and if they paid any attention to nonresponse bias. We found that seven of the studies did not use a mailed survey even though our initial search selected them. Searches such as this are only approximations because authors use the words in different contexts. For example, if an author wrote, “what is needed is a survey of the literature,” it would be included incorrectly in the search results. These two filters left seven articles in the final sample. Three did not mention the possibility of a nonresponse bias. Four addressed the issue in some fashion. One merely stated, “A high nonresponse is to be noted along with possible resulting bias” (Turner & Daly, 1965, p. 491). Another study reported the lack of a nonresponse bias and cited an unpublished reference to support these claims (Liou & Lawrenz, 2011).

We also searched the journal for articles that used web-based survey. We found two studies; however, neither of them used the web as a means of gathering survey data. One was an archival analysis of questions asked of scientists using a web-based service and the second was a book review that included the word “web.”

We found a similar situation for the *Journal of Research in Science Teaching* (JRST). A search using the terms “mailed (survey or questionnaire)” yielded 353 articles published between 1963, the first year of the journal, and 2011.

When we added the terms “nonresponse” and “nonrespondents” to the search, the number of hits dropped from 353 to 35. Each of these 35 articles was examined to determine its fit with the search criteria. This process excluded 13 articles because the surveys were not mailed or the word “survey” appeared in a different context. Of the remaining 22, only nine addressed the nonresponse issue. Three of these reported a nonresponse bias. For example, Gullickson (1978) used a telephone follow-up of nonrespondents in his survey of the familiarity of new elementary science curricula and found the responses were positively biased. Those familiar with the curricula were more likely to respond to his survey. Lawrenz (1990) surveyed teachers in the state of Minnesota and although she received an 86.3% response rate, she carried out a nonresponse study. She compared the demographic variables of her respondent and nonrespondent groups and found no differences between them.

We also searched issues of JRST regarding the use of web-based surveys or questionnaires. Only two were found. One was a survey of chemistry faculty members at higher education institutions in four states. The authors surveyed 400 instructors but only 13% responded. There was no mention of nonresponse bias in the article. The second web-based survey was an evaluation of a large NSF-funded teacher education program. Twelve of 19 projects agreed to participate for a response rate of 61%. Again, there was no mention of possible nonresponse bias.

We did find several references to nonresponse bias in the fields of agricultural education and university extension programs. For example, Lindner, Murphy, & Briers (2001) found that 53% of the articles published in the *Journal of Agricultural Education* between 1990 and 1999 mentioned nonresponse. The most common approach, about two-thirds, was to compare early with late responders. About a fifth of the studies followed up with a sample of nonrespondents.

A few studies compared respondents with nonrespondents and respondents with population background characteristics.

Although the use of late responders as a means of handling nonresponse issues in educational research dates back to Pace (1939) and was included in a text on questionnaire design (Oppenheim A. N., 1966), it is an article by Miller & Smith (1983) in the *Journal of Extension* that is quoted most often. They state, “Research has shown that late responders are often similar to nonrespondents” (p. 48). This way of addressing nonresponse issues has been readily heeded in the journals mentioned above and others similar to them. According to Google Scholar, authors have cited this article 508 times.⁵ In addition to the Miller and Smith claims, a textbook on questionnaire design (Oppenheim, 1966) states without reference, “it has been found that respondents who send in their questionnaire very late are roughly similar to nonrespondents” (p. 34). This approach has been used extensively in agricultural and extension education research but not in other education fields such as evaluation, science education, and technology education.

We decided to use this approach to investigate possible nonresponse bias with our survey but faced the problem of how to define a “late” responder. The only recommendation we found was one presented by Lindner, Murphy, Briers (2001, p. 52). They argue that late respondents should be those who respond to the last wave of contact. If the last stimulus does not generate 30 or more responses then use the last two contacts as the definition of late responders. If this does not yield 30 responses, they suggest using the later 50% of responders as the late responder group.

They also suggest, as do Borg and Gall (1983) that an alternative method would be to go back to the nonrespondents group, randomly sample at least 20 of them, and make an additional request for responses. We found two studies in *AJE* that did just that (Krushat & Molnar, 1993) and (Jones & Worthen, 1999). Both studies selected a random sample of nonresponders and attempted to contact them by telephone and mail. However, this turned out to be a difficult task and only about 50% of them actually replied to the telephone questions. This yielded another possible source of bias.

We did use this approach in our study but think it has merit. For those who wish to try this method, we recommend the article mentioned above for a description of the process (Gullickson, 1978). We now turn to our study of technological education grants to community colleges to determine potential nonresponse bias in our survey findings.

Method

A 23-item Likert scale survey was mailed to all active ATE grantees that began their grant prior to Jan 1, 2009 and to those grants that expired between Jan. 1, 2007 and Dec. 31, 2009. The survey consisted of statements made by fellow ATE PIs about the sustainability of their work. We asked respondents to agree or disagree with the statements made by their peers.

⁵ Google search April 2012.

We called this approach a Peer-Generated Likert Type survey. This is an effective way to generate survey items. One develops a description of the information that needs to be obtained. This is called the domain of content for the survey. People who have been involved in the object that is to be researched or evaluated are asked to make statements about their experiences. The process generates many statements that can then be used as items on the survey. They need to be reviewed and the most valid ones are selected for the survey. A detailed description of this process is found in a report on ATE sustainability (Welch, 2011). Here are examples of statements we obtained from ATE stakeholders.

- “Changes made in our technological education program will keep going after our current grant ends.”
- “We have at least one industry partner who is committed to support some of our work after NSF funding has ended.”
- “The instrumentation and equipment we secured as part of our ATE grant will have little use by the college after the grant ends.”
- “The teaching methods adapted by faculty as part of our ATE project will continue to be used after the grant ends.”

We used a five-point Likert format that ranged from Strongly Agree to Strongly Disagree. There was an option to mark “not applicable” if their grant did not do the activity mentioned in the statement. We coded these responses as missing values and excluded them from this analysis. We recoded negatively stated items so that disagree or strongly disagree answers indicated affirmation of sustainability. We added the responses for each subject and computed a reliability coefficient using Cronbach’s alpha. The value was .89 just below excellent level of .90 recommended by Mallery & Mallery (2003).

We used a number of procedures believed to enhance the likelihood of a subject responding to our request (Borg & Gall, 1983; Dillman D. A., 2000). We printed the survey in an attractive and colorful booklet, mailed it from a university, used a cover letter on NSF letterhead, offered a modest financial gratuity, and included a stamped, return-addressed envelope. Also, we used a new style of survey called Peer-generated Likert items that requests respondents to agree or disagree with statements made by their peers, offered to send them a copy of our findings, and most important of all, did three follow-up contacts. Thus, we had four waves of returns during our data collection period.

Data were collected between January and April 2010. The initial mailing included the survey, a cover letter, a return-addressed and stamped envelope, and a \$5 gratuity. Approximately two weeks later, we mailed a reminder letter to those who had not returned their surveys and two weeks later, we sent an email reminder message. Finally, in March, a complete packet, minus the \$5, was re-sent to all sites that had not responded by that date. We recorded the postmark date of each returned survey. This step is essential for researchers planning to use the early/late comparison to address nonresponse issues. It is how one identifies early and late responders.

Findings

Response Rates

One way to reduce response bias is to obtain a high percentage of returns. We address this issue here by reporting our response rate and comparing it to generally accepted standards. The returns for each wave are shown in Table 1.

Here one can see the benefits of repeated follow-ups. Although it is costly and time consuming, each wave yielded a substantial number of additional surveys. Notice that our three follow-up waves increased our response rate from 48% to 81%. This is probably the best procedure to use to ensure acceptable rates and should be at the forefront of consideration when planning an ATE research or evaluation study.

Table 1
Surveys Returned by Wave

Wave	Activity	Number Returned	Percent of Total Returned	Percent of Target Population (261)
One	Mailed First Packet	101	48%	39%
Two	Reminder Letter	48	23%	18%
Three	Reminder Email	31	14%	12%
Four	Re-Mail Packet	32	15%	12%
	Total	212	100%	81%

We reviewed the literature on survey response to discover what others are recommending. Fowler (1993) suggested 75% as a rule of thumb for minimum response. On the other hand, Babbie (1990) states that 50% is adequate, 60% is good and 75% is very good. Ary, Jacobs, & Razavieh (1996) argue that if the response rate is less than 75%, one should attempt to compare the characteristics of responders and nonresponders. One of the most widely used research methods books in education (Borg & Gall, 1983) recommends that if the return rate is less than 80%, one should check for possible nonresponse bias.

We reviewed articles in JRST and SCI EDUC that used mailed surveys to determine reported response rates. The average of 18 articles in JRST, that reported a response rate, was 68.9%. The average response rate for the articles we identified in SCI EDUC was 65.5%

We examined response rates in other fields as well. Baruch (1999) analyzed the mailed survey response rates in five journals published in the managerial and behavioral sciences. He found an overall average response rate of 55%. The response rate for top managers or organizational representatives was 38%. In addition, over the 20 years covered by the article, Baruch found the average response rate dropped from 64% to 48%. This finding is similar to the situation in other fields where response rates have dropped in recent years (Sheehan, 2001; Kaplowitz, Hadlock, & Levine, 2004).

We checked for articles that mentioned response rates to mailed surveys published in the American Journal of Evaluation and its predecessor, Evaluation Practice. We used the same period, 1963 to 2011, as we did for the science education journals. We found 31 articles that used surveys and 17 of them reported response rates. They ranged from 26% to 92%. The median return rate was 59%.

Our return rate compares favorably with figures recommended in research texts and exceeds the averages one finds in science education, evaluation, and the behavioral sciences. Although we met generally accepted standards for return rates, we carried out three studies to check for nonresponse bias. We compared the background characteristics of responders with the population and responders with nonresponders. We also compared the actual survey responses of early and late responders assuming that late responders were a reasonable surrogate for nonresponders.

Comparing Respondents to the Population

We obtained background information on our grantees from NSF's FastLane summary of awards. These included the following.

- Size of Grant: Average amount awarded in dollars
- Age of Grant: Average number of months between initial award and survey date
- Program Track: Grants were awarded for projects or centers. We used the percent of awards made for projects as our comparison variable.
- Institution Type: Two-year college, four-year college or other. We used the percent of awards made to two-year institutions for our comparison.
- Grant Status: Whether a grant was active or expired. We used the percent of active grants as our dependent variable.

We compared the responder group ($n = 212$) with the total population ($n = 261$) on the background traits listed above. We compared the means of the two groups for age and size of grant and used percentages for program track (percent projects), institution type (percent two-year colleges), and grant status (percent active).

The mean differences for the background characteristics are reported and expressed as effect sizes (ES). Effect size is a name given to a set of indices that measure the magnitude of differences between two groups. They often are used to compare a treatment group with a control group as part of meta-analysis studies.

We expressed our effect sizes using a statistic called Cohen's d . Cohen defined d as the difference between means, divided by the standard deviation of either group (Cohen, 1988). He also defined small ($d = .20$), medium ($d = .50$), and large ($d = .80$) effect sizes. Effect sizes can range from zero to 2.0 or higher. An effect size of 1.5 indicates a mean difference is 1.5 standard deviations. However, it is rare to find an ES above 1.0.

Effect sizes may be negative if the control group scores higher than a treatment group. In our case we are not actually comparing treatment and control groups; we are comparing responders with another group. We might consider the responders as our group of interest and the other groups as a comparison or control group. If the comparison groups, in the first case, the total population, scores higher than the responders, our effect size will be negative.

We show the results of our comparisons in Table 2.

Table 2
Comparing Responders with the Population on Five Background Characteristics

Characteristic	Responders Value ^a (212)	Population Value (261) ^b	Difference	Effect Size ^c
Mean Grant Size	\$826,700	\$793,168	\$33,532	.05
Mean Grant Age	46.3 mos.	47.0 mos.	-.7 mos.	-.03
Percent Projects	80.2%	82.4%	- 2.2%	-.03
Percent Two-Year Institutions	73.6%	73.9%	- .03%	.01
Percent Active Grants	61.8%	61.3%	.05%	.02

a Grant size and age are means. The other characteristics are percentages

b Sample size in the parentheses

c Expressed as Cohen's D

None of the comparisons between the 212 responders and the 261 grants in the population showed any differences. All ES were .05 or less which is very small. As mentioned earlier, this comparison is not very useful if there is a high response rate. This is because there is a large overlap between the two groups because the 212 responders are part of the total population. This method would be more appropriate when response rates are low, say less than .60. This would reduce the amount of overlap and increase the probability of detecting differences if they existed.

Comparing Respondents with Nonrespondents

This approach compares the background traits of the responder and nonresponder group. We compared the two groups on size of grant, age of grant, the percent of projects, 2-year institutions, and active grants. The results of these comparisons are shown in Table 3.

Table 3

Comparing Responders with Nonresponders on Five Background Characteristics

Characteristic	Responders Value ^a (212)	Nonresponders Value (49) ^b	Difference	Effect Size ^c
Size of Grant	\$826,700	\$648,091	\$178,609	.28
Age of Grant	46.3 mos.	50.1 mos.	-3.8 mos.	-.17
Percent Projects	80.2%	91.8%	- 11.6%	- .31
Percent Two-Year Institutions	73.6%	75.5%	- 1.9%	.02
Percent Active Grants	61.8%	59.2%	2.6%	.06

a Grant size and age are means. The other characteristics are percentages

b Sample size in the parentheses

c Expressed as Cohen's D

There were no differences between the groups on the percent of active grants or two-year institution. Small effect sizes were found for the age of grant and for the size and the percent of projects. Older grants were slightly less likely to respond to the survey but the effect size was small.

In addition, smaller grants and the percent of projects as compared to the larger centers were less likely to respond. These two traits are highly correlated because centers are usually funded at higher levels than are projects. This confounds the analysis.

We regressed the size of grant and the percent project on the response/nonresponse variable. The R^2 value was small, .015, meaning that the two variables were not accounting for much variance in response rates. The beta weight for size was .03 and .10 for the percent projects. This means the project/center difference was the major factor in accounting for the small variance.

The low effect sizes indicate there is little nonresponse bias using these two methods. We now turn to the third approach we used, comparing early responders with late responders using their actual responses to our survey items.

Comparing Early Responders with Late Responders

This method of addressing nonresponse bias is based on the claim of several authors, for example, Lindner, Murphy, & Briers (2001), that late responders are similar to nonresponders. This allows one to use the late responder group as a surrogate for nonresponders. We made the

early/late comparison using two scale scores derived from the survey items. These were the ATE Sustainability Scale and the ATE Impact Scale.

The Sustainability Scale includes 23 Likert-type items that are statements about sustainability made by ATE project leaders and other stakeholders. Respondents were asked to agree or disagree with these statements using a 5-point scale that goes from Strongly Agree to Strongly Disagree. Scale scores were obtained by summing the responses. The reliability of the scale using Cronbach's alpha was .89. Here is an example of one of the items "Changes made in our technological education program will keep going after our current grant ends." The development and analysis of the scale is described in Welch (2012).

The Impact Scale contained 29 Likert-type items that were statements about the impact of the ATE experience as described by principal investigators and people knowledgeable about the program. The response options were the same as those described in the preceding paragraph. The reliability of the scale was .88. A sample item is "Because of our grant, local industries are more willing to ask us to provide technological education for their workforce." Details about the Impact Scale and its psychometric properties can be found in Welch (2013).

This way of addressing possible response bias compares early and late responders. However, it is not immediately clear how to define a late responder. Miller and Smith (1983) used the term "late" without further explanation. Oppenheim (1992) mentioned very late responders without defining what "very" late means. However, we found a recommendation in Lindner, et.al, (2001) that seems reasonable. They argue that late respondents should be those who respond to the last of several waves of contacts. If the last stimulus does not generate 30 or more responses then one should use the last two stimuli as the definition of late responders. If this does not yield 30 responses, they suggest using the later 50% of responders as the late responder group. The authors do not explain their choice of 30 as a required minimum sample size but it is consistent with the recommendations found in educational research textbooks. For example, Borg and Gall (1983) suggest a minimum of 30 subjects for correlational studies and 15 per group for experimental research (p. 257).

Our wave 4 contact resulted in 32 responses (see Table 1) so we decided to use that group as our "late" responders. Wave 1 yielded 101 returned surveys, which became our "early" responders. We compared these two groups using the scale scores described above. Our findings are shown in Table 4.

Table 4

Comparison of Early and Late Responders on Two ATE Scales

	Wave 1 & Wave 4	N	Mean	S.D.	Difference	Effect Size
Impact Scores	First Responders	101	92.50	15.68		
	Last Responders	32	92.41	17.55	.09	.01
Sustainability Score	First Responders	101	74.50	17.20		
	Last Responders	32	73.13	16.72	1.37	.08

We found no significant difference between the early responders and late responders on either scale. The effect sizes were negligible. By inference, we conclude there are no differences between responders and nonresponders. Evidently, the slight response bias we found toward larger centers when examining background characteristics did not transfer to the actual responses to the statements in the sustainability and impact scales. The three methods we used lead us to believe that we did not have a nonresponse bias on our surveys.

Concluding Remarks

The general purpose of this report is to raise awareness of potential problems and solutions caused by survey nonresponse. We summarized the research on nonresponse issues, offered suggestions on how to increase response rates, presented generally acceptable standards for response rates, described methods to check for nonresponse bias, and applied these procedures to a research study of the Advanced Technological Education (ATE) program sponsored by the National Science Foundation (NSF).

We described various techniques one can use to increase responses to mailed surveys and pointed out that a very effective approach is to carry out follow-up surveys. We used three follow-up waves and increased our original response rate from 48% to 81%.

We explored the literature regarding acceptable response rates and found most texts recommend 70-80%. Not all journal articles report their response rates but when provided, generally they are in the 50-65% range. The range for the articles we reviewed was 26% to 100%.

We described a data gathering procedure that yielded an 81.2% response rate even though the respondents were unfamiliar with us or with our study. Techniques we used to enhance response rates included the following. Respondents were asked to respond to statements written by their peers. The survey was printed as a colorful 4-page booklet that took about 20 minutes to answer. It included a cover letter with a NSF logo and we provided a stamped return-addressed envelope to make the process easier for our respondents. We included a small stipend of \$5 and offered to send returnees a copy of our final report. Most important of all, we sent out three follow-up reminders.

We used three methods to check for potential nonresponse bias for those that did not return the survey. We compared the background traits of responders and the total population and presented the results as effect sizes. All effect sizes were nearly zero indicating no response bias. However, part of this can be explained because of the high overlap between the groups. The population of 261 grants included the 212 responders. This method would be more effective when the response rate is lower. This would reduce the amount of overlap between the two groups.

The second approach was to compare the background characteristics of the 212 responders with the 49 nonresponders. There is no overlap of these two groups. We found that centers and/or higher-funded grants were more likely to respond to our survey. These two issues were related because centers, on average, receive more NSF support than projects. We used

multiple regression analysis to determine the relative contribution of size and the project/center difference. All the effect sizes are small, this analysis suggested that it is the project/center difference that explains more of the variance than does size. Most centers are quite stable and are funded year after year. We suspect there is less staff turnover and the institutional memory is greater.

We also tested for bias using the assertion that late respondents are a reasonable surrogate for nonrespondents. We compared the survey responses for the early and the late groups. Our late group consisted of the 32 sites who responded to our fourth contact. The early group consisted of the 101 sites that responded to the first mailing. We found no significant difference ($\alpha = .05$) between the early/late scores on either the sustainability or the impact scale. By inference, we concluded there was no difference between those that returned the survey and those that did not. The background bias we found for the project/center variable did not seem to carry over to how people replied to the survey questions.

Of course, there is the question of whether or not late responders are a reasonable surrogate for nonresponders. Pace (1939), one of the first to use the method, conducted research on the matter using questionnaires sent to graduates of universities. He concluded, "It provides a simple and valuable tool for determining the probable direction of bias" (p. 397). Miller and Smith (1983) are more direct, "Research has shown the late responders are often similar to nonrespondents" (p. 48). Lindner, Murphy, & Briers (2001) point out the prevalence of the method in other educational fields. However, caution is urged when using this approach. As Oppenheim (1966) stated, the early vs late comparison can be used as a rough estimate of nonresponse bias. It helps identify potential sources of error but does not indicate the extent of bias.

Nonresponse is an ongoing problem in survey research. We believe it is important for PIs and others conducting ATE research and evaluation to heed the suggestions provided in this report to help ensure the validity of their survey findings.

References

- Ary, D., Jacobs, L., & Razavieh, A. (1996). *Introduction to research in education*. Fort Worth, TX: Holt, Rinehart, and Winston.
- Babbie, E. R. (1990). *Survey research methods*. Belmont, CA: Wadsworth.
- Baruch, Y. (1999). Response rates in academic studies-A comparative analysis. *Human Relations*, 451-468.
- Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction*. New York and London: Longman.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd. ed.)*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Converse, P. D., Wolfe, E. W., Huang, X., & Oswald, F. L. (2008). Response rates for mixed-mode surveys using mail and e-mail/web. *American Journal of Evaluation*, 99-107.
- Coryn, C. L., Ritchie, L. A., & Gullickson, A. R. (2006). *ATE Evaluation Fact Sheet*. Kalamazoo, MI: Evaluation Center, Western Michigan University.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method (2nd ed.)*. New York: John Wiley.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail and mixed-mode surveys: The tailored design method, (5th ed.)*. Hoboken, NJ: John Wiley.
- Flowers, J. (2001). Online learning needs in technology education. *Journal of Technology Education*, Volume 13, Number 1.
- Fowler, F. J. (1993). *Survey research methods (2nd ed.)*. Thousand Oaks, CA: Sage.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley.
- Gullickson, A. R. (1978). The impact of a mass media approach to disseminate elementary science curricula. *Journal of Research in Science Teaching*, 325-333.
- Hartge, P. (1999). Raising response rates: Getting to yes. *Epidemiology*, 105-107.
- Jones, S. C., & Worthen, B. R. (1999). AEA members opinions concerning evaluator certification. *American Journal of Evaluation*, 495-506.
- Kaplowitz, M. D., Hadlock, T. D., & Levine, &. a. (2004). A comparison of web and mail survey response rates. *Public Opinion Quarterly*, 94-101.
- Kiernan, N. E., Kiernan, M., Oyler, M. A., & Giles, C. (2005). Is a web survey as effective as a mail survey? A field experiment among computer users. *American Journal of Evaluation*, 245-252.
- Krushat, W. M., & Molnar, J. I. (1993). The effect of nonrespondents on self-administered mail survey. *Evaluation Practice*, 25-32.
- Lawrenz, F. (1990). Science teaching effects associated with higher-order thinking skills. *Journal of Research in Science Teaching*, 835-847.
- Lindner, J. R., Murphy, T. H., & Briers, G. E. (2001). Handling nonresponse in social science research. *Journal of Agricultural Education*, 43-53.
- Liou, P., & Lawrenz, F. (2011). Optimizing teacher preparation loan forgiveness programs: Variables related to perceived influence. *Science Education*, 121-144.
- Mallery, G. M., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference. 11.0 update (4th ed.)*. Boston: Allyn and Bacon.
- Manfreda, K. L. (2008). Web surveys sersus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research*, 79-104.

- Miller, K. L., & Smith, L. E. (1983). Handling nonresponse issues. *Journal of Extension*, 45-50.
- Odom, A. L., Settlage, J., & Pederson, J. E. (2002). Technology knowledge and use: A survey of science educators. *Journal of Science Education and Technology*, 391-398.
- Oppenheim, A. N. (1966). *Questionnaire Design and Attitude Measurement*. New York: Basic Books, Inc.
- Pace, C. R. (1939). Factors influencing questionnaire returns from former university students. *Journal of Applied Psychology*, 388-397.
- Sheehan, K. (2001). E-mail survey response rates: A review. *Journal of Computer Mediated Communication*, January.
- Singer, E. (2006). Introduction: Nonresponse bias in household surveys. *Public Opinion Quarterly*, 637-645.
- Turner, N. D., & Daly, S. S. (1965). Attitudes of upstate New York section students toward the maa-sa contest. *Science Education*, 491-493.
- Welch, W. W. (2012, June). *Measuring the sustainability of the advanced technological education (ATE) program*. Retrieved from Evalu-Ate: http://evaluate.org/resources/sustainability_of_ate/
- Welch, W. W. (2013, January). *Developing and analyzing a scale to measure the impact of the advanced technological education program*. Retrieved from Evalu_ATE: http://evaluate.org/featured_resources/resources/ate_impact_report_welch_2013/
- Welch, W. W. (Revised December 2011). *The sustainability of the advanced technological education program*. Retrieved from Evalu_ATE: <http://www.colorado.edu/ibs/decaproject/pubs/index.html>