Can We Count on Counting? An Analysis of the Validity of Community Engagement Survey Measures

Ethan A. Kolek

Can We Count on Counting?
An Analysis of the Validity of Community Engagement Survey Measures

Ethan A. Kolek
Amherst College

Researchers use various survey efforts to understand students' community engagement experiences. Among the crucial pieces of information for both academic and applied research is the extent to which (or whether or not) students participate in community engagement activities. However, recent studies have questioned the validity of many college student survey items. This paper describes an exploratory study that sought to investigate the validity of several survey items related to students' community engagement participation. The study found that large percentages of students who have taken community-based learning courses do not accurately report these experiences on student surveys and examines what factors relate to misreporting. Implications for future community engagement research are discussed.

Keywords: surveys; measurement error; community engagement; service-learning

Over the past two decades, there has been a dramatic increase in the number of civic and community engagement efforts at American colleges and universities (Bringle, 2011). Currently, over 1,100 higher education institutions are members of Campus Compact, a national coalition fostering civic engagement (Campus Compact, n.d.), and the prevalence of the many forms of service-learning and civic engagement activities continues to grow (Hatcher & Erasmus, 2008). As service-learning and civic engagement programs have become ubiquitous at colleges and universities, higher education researchers have sought to answer a number of questions about these efforts. For example, one strand of research seeks to determine if participation in service-learning courses, community service, volunteer work, or other community engagement activities is related to positive educational outcomes, such as academic development and sense of civic responsibility and social justice (e.g., Astin & Sax, 1998; Bowman, Brandenberger, Mick, & Smedley, 2010). Another examines the factors that seem to influence students' involvement in community engagement or volunteer activities (e.g., Chesbrough, 2011; Cruce & Moore, 2007). A third strand uses measures of volunteering, service-learning, and community service as independent variables in studies examining other important student outcomes, such as leadership development (e.g., Dugan & Komives, 2010). Like other branches of education literature (Hutchinson & Lovell, 2004), many of these studies rely on data collected from student surveys (e.g., Astin & Sax, 1998; Bowman et al., 2010; Cruce & Moore, 2007; Dugan & Komives, 2010; Hurtado, Enberg, Ponjuan, & Landreman, 2002; Marks & Jones, 2004).

In addition to these scholarly studies, several prominent surveys used by colleges and universities for assessment, evaluation, planning, and policy decisions include items that ask students to report about their volunteering activities or community service, including the National Survey of Student Engagement (NSSE) and the Higher Education Research Institute's (HERI's) senior survey. In fact, two of the five NSSE benchmarks of effective educational practice, *Active and Collaborative Learning* and *Enriching Educational Experiences*, include items asking students about their participation in community engagement activities (i.e., participating in "a community based project as part of a regular course" [p.1], and "participating in… community service or volunteer work"

[p.2]) (National Survey of Student Engagement, n.d.). Campus leaders can use these measures to understand the extent to which students participate in community engagement activities, what groups of students are more likely to participate than others, and how participation relates to other educational experiences and outcomes. These data are used to assist in the evaluation of campus programs and to assess outcomes potentially associated with these engagement activities. Among the crucial pieces of information for both academic and applied research on community engagement is the extent to which (or whether or not) a student has participated in community engagement activities. Results from these studies can play an important role in high-stakes decision-making processes, particularly given current concerns for institutional accountability. Because these survey results have real consequences for our understanding of higher education from researcher, practitioner, and policy-maker perspectives, it is important to ensure that community engagement survey items are adequately capturing what they purport to measure.

This study is an extension of an evaluation project that I conducted at one community engagement center. In the course of analyzing survey data, I found that estimates of some community engagement behaviors, when examined by class year, did not seem to match corresponding administrative data. This discrepancy spurred me to investigate the validity of survey items that seek to measure students' participation in two common service-learning activities— participation in community-based learning courses and public service internships. Moreover, to the extent that these survey measures were producing invalid results, I sought to understand why. To provide context for this study, the next sections review the role of survey research in understanding the college student experience, as well as the survey research methods literature that focuses on the survey response process and item design. Although this study examines student survey items about community engagement activities specifically, the underlying concepts and potential problems are applicable to all student surveys in general (see Porter, 2011). Following the review of the survey methods literature, I describe the study's three data sources, community engagement measures, and data analysis. I then present and discuss the study's findings and acknowledge its limitations. The paper concludes by suggesting implications for community engagement and service-learning research and for college student surveys in general.

Surveys Research in Higher Education

Surveys are one of the most common ways to collect data for higher education research (Hutchinson & Lovell, 2004; Pike, 2008), and the majority of higher education quantitative studies focus on college students (Porter, 2011). Often, surveys about college students' participation in a variety of academic and co-curricular activities are used to understand the mechanisms of student learning, to assess engagement, or to evaluate programs (Porter, 2011). Theoretically, surveys are well-suited to these efforts because they allow researchers to capture data about a variety of domains, are relatively efficient data-collection methods (Krathwohl, 1998), and, when properly designed and well-executed, produce results that are generalizable to the population in question (Groves, Fowler, Couper, Lepkowski, Singer, & Tourengeau, 2009). Furthermore, higher education researchers often employ surveys because variables of interest—for example student attitudes or specific behaviors— are not available through administrative records or other institutional data (Porter, Rumann, & Pontius, 2011).

Of course, surveys are subject to several sources of error, most importantly nonresponse, sampling, coverage, and measurement errors (see Dillman, Smith, & Christian, 2009, and Groves et al., 2009, for excellent introductions to these concepts). The extent to which researchers can successfully estimate population parameters and understand interrelationships among survey variables of interest is contingent upon obtaining valid survey data (Groves et al., 2009). Because coverage error and sampling error can often be minimized in surveys executed within institutional

settings (Jans & Roman, 2007), nonresponse error and measurement error are often the most potentially problematic sources of error in surveys of college students.

Unfortunately, the validity of many items used in surveys of college students has not been established (Porter, 2011). Historically, the higher education research literature has not devoted much attention to survey methodology, in part due to the lack of methods training in graduate programs (Hutchinson & Lovell, 2004; Malaney, 2002; Porter, 2011). Recently, several studies have questioned the validity of a number of commonly used student survey measures related to learning, grades, and students' academic behaviors (Bowman, 2011a, 2011b; Bowman & Brandenberger, 2010; Kuncel, Crede, & Thomas, 2005; Porter, 2011; Porter et al., 2011). In fact, Porter (2011) argued that the field of higher education "requires an ambitious research program to reestablish the foundation of quantitative research on students" (p. 45). In response to Porter's (2011) critique of higher education's survey enterprise in general, and the National Survey of Student Engagement (NSSE) in particular, McCormick and McClenney (2012) pointed to the lack of empirical evidence that college students respond "incorrectly" to survey questions. This study seeks to contribute to this conversation by offering some empirical evidence about the validity of students' survey responses in the community engagement domain focusing on potential measurement error.

Survey Methodology

Survey Response Processes

Obtaining valid and reliable survey data is a difficult task, in part because the survey response process is complicated, requiring survey researchers to take into account the limits of human comprehension, memory, and linguistic nuances when designing surveys (Dillman et al., 2009; Foddy, 1993; Fowler, 1995; Groves et al, 2009; Sudman & Bradburn, 1982; Tanur, 1994). Tourengeau, Rips, and Rapinski (2000) outlined a four-step process describing how individuals respond to survey questions. Potential respondents must first understand the question being asked, and all potential respondents should be able to understand the question to mean the same thing. Then the respondent must retrieve the relevant information from memory, judge or evaluate the information to determine whether it does indeed match the question, and finally respond. As Groves et al. (2009) point out, respondents may move back to any previous step in order to answer the question. In practice, breakdowns can easily occur at any stage of this process, so it is imperative that researchers design items that minimize errors at each step.

Survey Effort

In addition to considering the implications of the four-stage response process on item design, survey researchers must consider respondents' motivations to answer questions and the effort respondents will be willing to exert to complete a survey. Krosnick and Alwin (Krosnick, 1999; Krosnick & Alwin, 1987) proposed a model of survey response whereby each respondent employs one of two parallel processes. In the first, survey respondents make efforts to provide the most accurate information possible, referred to as "optimizing." In the second path, respondents use a minimal amount of effort to complete the questionnaire, referred to as "satisficing." Respondents may satisfice rather than optimize for a number of reasons—for example, because they feel burdened by the length of a survey, because survey questions are too cognitively taxing, or because they are only grudgingly participating in the survey in the first place (Krosnick, 1999). Respondents who satisfice do not necessarily understand items completely, may not be exhaustive in their efforts to retrieve information, and may not be discerning in their judgments. Because satisficers exert less effort to complete a survey, they are less likely to give valid and reliable responses than optimizers (Krosnick, 1999). Therefore, survey researchers need to construct instruments that are likely to minimize the

chances that respondents will satisifice and maximize the chances that respondents optimize (Groves et al., 2009).

Survey Items and Measurement Error

Measurement error due to respondents' difficulty in answering questions can occur for a number of reasons. Willis (2005) argued that the everyday conversations in which people ask each other questions do not translate easily to surveys because our language is often inherently ambiguous. In all conversations, individuals engage in the creation of a shared interpretation of the meaning of utterances (Willis, 2005). In contrast, surveys, in general, and self-administered surveys, by definition, are inflexible and do not permit interactions between two parties that can lead to common grounding between respondent and researcher (Willis, 2005). Misinterpretations of question meaning can happen for several reasons, including overly complex grammatical structures and the use of vague concepts or unfamiliar terms (Groves et al., 2009). The survey research literature provides numerous examples of seemingly simple questions that can cause measurement error due to differences in interpretation. For example, Fowler (1992) described an item asking people to report about their frequency of engaging in "exercise, sports, or physically active hobbies" (p. 222). In item testing, some respondents included walking in their reports, whereas others did not, necessitating revisions to the item to ensure that all respondents interpreted the question the same way. In the higher education context, Porter (2011) pointed out the ambiguities inherent in the phrase "critical thinking" and suggested that higher education researchers incorporate ill-defined jargon into survey items. Valid community engagement measures may be particularly difficult to construct given the wide range of language used to describe civic engagement/service-learning work (Saltmarsh, 2005). For example, more than a decade ago, Eyler and Giles (1999) pointed to nearly 150 different definitions of service-learning.

Another important aspect of item design is to avoid double-barreled questions—that is, items that ask respondents to report about two different things in the same question. Such items cannot produce valid data since some respondents will have divergent opinions but are forced into a single response choice. For example, in the question "How satisfied are you with the academic and social experiences at your institution?" some respondents may be very satisfied with their academic experiences and dissatisfied with their social experiences. Similarly, unless response options are of the "check all that apply" type, they must be mutually exclusive—a "cardinal rule for question writing" (Sheatsley, 1983, p. 217). For instance, as a forced-choice item, the question "Which of the following have you consulted for information about graduate school—professors, parents, classmates, or the career center?" is problematic because a respondent could have consulted more than one of these sources.

Another challenge to collecting accurate survey data relates to memory (Groves et al., 2009). Recall problems may be caused by several factors, including a mismatch between question language and language used to encode memories, and retrieval failure. People are more likely to recall specific events that have particular significance or salience than everyday events. Not surprisingly, memory research has found that more recent events are easier to recall than events further in the past (Foddy, 1993; Fowler, 1995). Additionally, recall problems can happen because of a respondent's inability to distinguish the particular details of one event among many similar experiences, what Groves et al. (2009) refer to as "generic memories"—amalgamations of similar events over time. However, particularly salient events, such as the first day of school, weddings, or personal tragedies, may be recalled with detail despite the passage of time (Fowler, 1995.)

Currently, we understand the survey response process to be quite complex, suggesting the need for great care in survey construction and examinations of survey validity. The survey methods literature shows that researchers need to compose surveys that encourage respondents to optimize rather than satisfice. To avoid problems in the survey response process, items need to be clear and

unambiguous so they can be comprehended by all respondents. They should only ask for information that would be reasonable for respondents to have encoded and are able to retrieve, and they should be written in a format to allow for valid responses. Unfortunately, the validity of many higher education survey items is unknown, and some studies suggest that higher education surveys do not produce valid results (Porter, 2011). Community engagement measures are not immune to this criticism. Given the importance of these types of data for research and policy decisions, community engagement researchers need to be sure that their measures are rigorously evaluated. To that end, this study seeks to inform researchers' understanding of the validity of several community engagement survey items.

Methods

This study uses data collected at a single, highly selective, private liberal arts college in the Northeastern United States. The college enrolls approximately 1,800 undergraduates, all full-time students, the vast majority (97%) of whom live on campus. Less than 1% of students are over the age of 24. This study draws from three data sources: (a) administrative data maintained by the college for the entire population of undergraduates, (b) a general student web survey sent to the population of undergraduates (part of a multi-institution effort), and (c) a local web survey about community engagement that was sent to the population of undergraduates. The general student (GS) survey was designed by a group of researchers from several different campuses, whereas the community engagement (CE) survey was designed for this particular campus. Data from these three sources were matched and combined in a single data file in order to compare students' self reports with the college's administrative records regarding their participation in two types of community engagement activities—community-based learning courses and public service internships. This study uses these combined data to answer the following questions: (a) to what extent do students' self-reports of participation in community-based learning courses and public service internships correspond to administrative records?; and (b) to the extent that survey responses do not match administrative data, what factors relate to misreporting?

Administrative Data Measures

Course data. Historically, the institution in this study has offered few service-learning courses, most commonly referred to as community-based learning (CBL) courses at this college. With the establishment of a new community engagement center in 2007, staff members conducted extensive outreach to faculty and engaged in faculty development work to understand the scope of CBL course offerings and faculty involvement in community engagement activities. As part of this effort, information about CBL courses was compiled in an institutional database. At the time of this analysis, the database contained CBL course listings and enrollments from fall 2007 through spring 2011 (eight semesters). The center had compiled a database of 39 CBL courses offered between fall 2007 and spring 2011 (606 enrollments from degree-seeking students). Because of the relatively small size of the institution and strong working relationships between center staff members and many faculty members (e.g., over a two year period, one staff member had meetings with approximately one-third of full-time faculty) the center's staff members were confident that the database captured the vast majority, if not all of the CBL courses that had been offered. Student enrollment data were extracted from the college's database for all students who were enrolled in one of these courses.

   Internship data. The community engagement center runs a competitive public-service internship program to which students apply to receive stipends to cover living expenses while engaging in summer internships for a minimum of 320 hours in the nonprofit or government sector. However, not all students who participate in public-service internships are part of this program. For

the purposes of this study, two elements of this program are particularly salient. First, because students received funding from the institution, a comprehensive tracking system is in place to capture information about all students who participated in this program, since these data are necessary to process stipends. Second, the words "intern" and "internship" were used in all publicity, application, contracts, and evaluation materials to describe these activities. Although in some contexts the term "internship" may be ambiguous, we would expect all participants in these programs to think of these summer experiences as "internships."

Survey Measures

In spring 2011, two web surveys were administered to the entire population of undergraduates. The GS survey was conducted under the auspices of the Office of Institutional Research and achieved a response rate of 60.5% (n=1,078). Students were sent an email invitation, and nonrespondents were sent up to two reminders asking them to complete the survey. The survey consisted of a minimum of 125 questions (respondents were asked additional questions if they had declared a major). The CE survey was sent under the auspices of the community engagement center and had a final response rate of 21.2% (n=378). Students were sent an email invitation, and nonrespondents were sent up to three reminders asking them to complete the survey. The CE survey had a minimum of 38 items asked of all respondents. For both surveys, participants were similar to the population in terms of class year. Women were overrepresented compared to men in both surveys. Women comprised 51% of the target population but 55% of respondents to the general survey and 60% of respondents to the CE survey. This is unsurprising as numerous studies have found that female college students are more likely to respond to surveys then male students (Pike, 2008; Porter &Whitcomb, 2005).

Two survey items from the GS survey and two items from the CE survey are the primary survey measures used in this analysis. The GS survey asked students to report whether they had done, planned to do, did not plan to do, or were undecided about participating in a service-learning course or program. The survey used the same response scale in asking students to report about having an internship during their time at the institution.

The CE survey asked students to report whether or not they had taken a course with a community-based learning component during the current academic year and, if they had done so, to report the name of the course or courses. Respondents were also asked whether or not they had a public-service internship during the previous summer. If respondents indicated that they had done so, they were asked if the college had funded it. In the CE survey, only sophomores, juniors, and seniors were asked about having a public- service internship because first-year students would not have been able to receive college funding during the previous summer.

Correspondence of Measures

For each survey respondent, responses to each item were matched to the relevant administrative data to ascertain the accuracy of students' self-reports (see Figure 1).

Using the example of a CBL course, block A corresponds to a case in which enrollment data and a student's survey response both indicated that a student had taken a CBL course (true positive). Block B corresponds to a case in which enrollment data indicated that a student had taken a CBL course but in which the student marked that he or she had not taken a CBL course on the survey (false negative). Because of the very high level of certainty that all enrollment data that show that a student had enrolled in a course are accurate (blocks A and B), we can examine true positives and false negatives for all survey measures. For example, if a case is marked as "enrolled" in administrative data, that student has received academic credit in a course that has been judged to be a community-based learning course by staff members of the engagement center. We need to exercise greater caution in interpreting the cases for which there is no enrollment data (blocks C and D). For

example, a student may have taken a course with a community-based learning component while studying abroad at another institution. In this instance there would be no enrollment record that she or he had completed a service-learning course. Although such cases are likely to be rare, it is important to note that we have less certainty that all cases in block C and block D are coded correctly.

| | Database Record | |
Survey Response	Enrolled	No Enrollment Record
Enrolled[1]	(A) True Positive	(C) False Positive
Not Enrolled[2]	(B) False Negative	(D) True Negative

[1] "Done" (GS Survey)/ "Yes" (CE Survey)
[2] "Have Not Done," "Plan to Do," or "Do Not Plan to Do" (GS Survey)/ "No" (CE Survey)

Figure 1. Classification of survey response based on corresponding administrative records.

Data Analysis

Frequency distributions were computed for administrative and survey data for CBL courses and public-service internships to examine the extent to which survey estimates corresponded with administrative records. Second, cross-tabulations were run to examine the extent to which respondents accurately coded themselves as having taken a CBL course and having participated in an internship. Third, a series of cross-tabulations and mean comparisons were run to explore potential differences between true positives (cases in which the survey response and administrative record both indicated that a student had engaged in a particular behavior) and false negatives (cases in which survey response indicated a student had not engaged in a behavior but where the administrative record indicated that the student had actually engaged in the behavior). I used several indicators to examine the extent to which survey effort and recall issues may be related to misreporting. The measures of survey effort were number of items with missing data, time to complete the survey (in minutes), and whether the student completed the survey after receiving the invitation, the first reminder, or the second reminder.[1] These measures have been used in several previous studies to gauge survey effort (e.g., Heerwegh & Looseveldt, 2008; Petrolia & Bhattacharjee, 2009; Porter &Whitcomb, 2003; Singer, Van Hoewyk, & Maher, 2000; Teisl, Roe, & Vayda, 2005).

I also examined several factors that might suggest students experienced recall problems. Students' class year was examined because students who had taken more courses may be less able to recall salient features of individual courses. For example, a senior may be attempting to retrieve data about one course out of over 30, whereas a first-year student may be attempting to retrieve information about one of eight courses. Second, because more recent activities are more easily recalled than activities that occurred further in the past, how recently a student enrolled in a CBL course was examined. Third, I examined the number of CBL courses in which students enrolled because taking more than one course should increase the chance that students could recall their CBL

course participation. Finally, it is likely that community-based learning experiences differed among courses; therefore, I examined whether students who took particular courses seemed to disproportionately misclassify themselves.

Results

Comparison of Survey Distributions to Administrative Data

Analysis of the CE survey revealed that 17% of students had participated in a service-learning course or program during their time at the institution (see Table 1). This is within two percentage points of the administrative data for the population, which indicated that 15% of students had taken a CBL course.[2] Similarly, the CE survey estimates of CBL course enrollment in the past year is within two points of administrative data records of all students. The CE survey produced an estimate of college-funded public-service internship participation within three percentage points of the administrative data. In terms of face validity, this survey item corresponded well to available administrative data.[3] These three items—again, in terms of face validity—seem to match the domains captured by the relevant administrative data, so it is not particularly surprising that the survey estimates of participation are close to those calculated from administrative data.

Table 1. Comparisons of Administrative Data Distributions to Survey Distributions

	GS Survey	Corresponding Enrollment Data (GS)	CE Survey	Corresponding Enrollment Data (CE)
CBL Course	17%	15%	12%	10%
No CBL Course	83%	85%	88%	90%
Internship[a]	[b]	[b]	13%	10%
No Internship[a]	[b]	[b]	87%	90%

[a] Internship percentages calculated for sophomores, juniors, and seniors.
[b] Enrollment data were only available for college-funded, public service internships.

Classification Errors

Next, I examined the extent to which students who had (a) taken a course with a community-based learning component and (b) completed the GS survey (n=181) or the CE survey (n=55) correctly classified their behavior on each instrument. As can be seen in Table 2, less than one-half of the students who took a course with a community-based learning component reported having done so on the GS survey. Respondents were relatively evenly split between marking that they planned to take such a course or program, did not plan to do so, or were undecided.

On the CE survey, two-thirds of the students who had enrolled in a community-based learning course during the current academic year correctly classified their behavior, whereas one-third incorrectly reported that they had not taken such a course (see Table 3).

Table 2. GS Survey: Survey Response for CBL and No CBL Students

Survey Response*	CBL Course Enrollment Record	No CBL Course Enrollment Record
Done	48%	10%
	(87)	(91)
Plan to do	15%	20%
	(27)	(170)
Do not plan to do	21%	39%
	(38)	(341)
Have not decided	16%	31%
	(29)	(270)
Total	100%	100%
	(181)	(872)

Note: Survey question: Which of the following have you done or do you plan to do during your time at [Institution X]: Participate in a service learning course or program?

Table 3. CE Survey: Survey Response for CBL and No CBL Students

Survey Response*	Enrolled in CBL Course: Current Academic Year	Not Enrolled in CBL Course: Current Academic Year
Yes	62%	3%
	(34)	(9)
No	38%	97%
	(21)	(292)
Total	100%	100%
	(55)	(301)

Note: Survey question: This academic year, have you taken a course with a community-based learning component vs. enrollment data?

Next, I examined the extent to which students who had completed a public-service internship (according to administrative data [N=205]) correctly classified their experience on each of the two surveys. Of these students, 140 responded to the GS survey. The vast majority (94%) correctly reported having completed an internship. All 27 of the students who participated in the CE survey and had a college-funded public-service internship in summer 2010 correctly reported having had such an internship. In contrast to the widespread error in reporting CBL course enrollment among participants, nearly all college-funded public-service interns correctly reported their participation.

Potential Sources of Error

Response effort. The next analysis compares cases of true positives (block A) and false negatives (block B) to the item "Have you participated in a service learning course or program?" on the GS survey. Students who were not enrolled in CBL courses according to administrative data (blocks C and D) are excluded from these analyses. Among those who completed the entire survey, true positives and false negatives did not meaningfully differ in the amount of time spent completing the survey, the number of items with missing data (out of the 112 common forced-choice items), or in the percentage of those who completed the survey after receiving the initial invitation, the first reminder, or the second reminder (see Table 4).

Table 4. GS Survey: Effort and Recall Comparisons for CBL Students

Domain	Measure		True Positive (A) (n=87)	False Negative (B) (n=94)
Effort				
	Mean minutes responding to survey		20.0	19.7
	Mean item nonresponse		1.5	2.5
	Respond to survey after:	Initial Invitation	46%	47%
		Reminder 1	30%	38%
		Reminder 2	24%	15%
Recall				
	Most recent CBL course	Same semester	52%	48%
		Previous semester	41%	59%
		Two semesters previous	52%	48%
		Three semesters previous	36%	64%
		Four or more semesters previous	72%	26%
	Number of CBL courses	One	45%	55%
		Two or more	63%	37%
	Class Year Distribution	First year	29%	71%
		Sophomore	53%	47%
		Junior	40%	60%
		Senior	60%	40%

Recall. Greater salience, as measured by taking one or more than one CBL course, was related to accuracy of reporting. Of the 149 students who had taken one CBL course, less than one-half correctly classified themselves compared to nearly two-thirds of the 32 students who had taken more than one CBL course. Surprisingly, there were no clear differences between these groups on the other two recall measures. Recency of enrolling in a CBL course had no clear relationship to respondents' accuracy of reporting. Examination of the role of class year produced unexpected results, with less than one-third of first year students correctly reporting that they had enrolled in a CBL course compared to between 40% and 60% of sophomores, juniors, and seniors.

Item misinterpretation. Exploratory analyses were conducted to ascertain if particular CBL courses disproportionately enrolled students who misclassified themselves as non-CBL students. Given the very small numbers of students who were currently enrolled at the college, had enrolled in a particular course, and completed the survey (mean = 6.1, median = 6), it was difficult to draw meaningful conclusions from much of these data. Of the 34 courses with at least one student who responded to the survey, seven had no students misclassify, and four had no students correctly classify themselves.

Looking at students who participated in one English course illustrates inconsistencies in reporting. In this course, which was offered eight times during the period covered by these data, students engaged in weekly tutoring or served as a classroom aide in a local high school or with an adult literacy project. According to enrollment data, about one-third of all students who had enrolled in a CBL course and who completed the survey (n=67) had enrolled in this particular course. Overall, more than one-fifth of these students (22%) did not report that they had participated in a service-learning course. Misreporting differed by semester, ranging from a low of 57% of students correctly reporting to a high of 100%.

The nine false positives from the CE survey provide some limited insight into other potential misinterpretations of the CBL course survey item. Five of the nine false positives reported having enrolled in CBL courses that center staff confirmed did not actually have a community-based learning component. One student reported that his or her thesis had a community-based learning component. Since the student database does not include detail about individual students' theses, it is certainly possible that this student's report was accurate. Finally, three respondents reported taking a two-week, non-credit, non-academic CBL course during the institution's January break. Although it was intended that students report about their academic coursework, these offerings are indeed referred to as "courses" at this institution.

Discussion

Students' responses to items asking about service-learning or CBL course participation differed quite substantially from enrollment data, and it appears that widespread error exists for these survey questions. On the GS survey, one-half of CBL course participants incorrectly reported that they had not taken a course with a community-based learning component, and on the CE survey over one-third of such students incorrectly reported that they had not enrolled in such a course. In contrast, in both surveys the vast majority of students who had participated in a public-service internship correctly reported having done so.

One common heuristic some researchers use to judge the accuracy of survey data is to compare a survey response distribution to a known population distribution (e.g., demographics characteristics or political party membership). If respondents match the population with regards to these characteristics, researchers believe there is a lower chance of biased results. These data reveal the potential problems of comparing survey response distributions to population characteristics in order to determine the accuracy of a survey measure. Response distributions to some survey items, most notably the general survey item asking about participation in service-learning courses or programs, did not differ from administrative data despite substantial misreporting by respondents. This is due to the presence of false positives (block C) who reported they had taken a CBL course but for whom there was no enrollment record that they had done so. Because this group was similar in size to the false negative group, it seems that their combined misreporting appeared to cancel the other out when calculating the overall percentage of students who had enrolled in a CBL course, unfortunately obscuring the error.

Response Category Explanation for Misclassification

One potential problem with the GS survey items is that the response categories are not mutually exclusive, yet the questions are designed to be forced choice. In other words, a student may have "done" an internship as well as "plan to do" an internship in the future. There are no instructions indicating that a student should only mark "plan to do, do not plan to do, or have not decided" if "done" does not apply. So, it is possible that some students selected the first option then proceeded to mark the next valid option without realizing that in doing so their first response was not saved. This item violates a basic element of survey question design by asking multiple questions in a single item (Sheatsley, 1983). Although these exact items may be unique to this particular survey, this question format has been used in other prominent higher education surveys, such as the 2011 NSSE. However, if the response option format was the main cause of misreporting about service-learning courses, there would likely be similar misreporting about internships, suggesting that the response option format did not cause substantial misreporting.

Satisficing Explanation of Misclassification

None of the three measures of respondent effort suggested that misclassification was due to satisficing. However, given the climate of "survey fatigue" that several researchers have identified on college campuses (Asiu, Antons, & Fultz, 1998; Porter, 2005; Porter, Whitcomb, & Weitzer, 2004) and the number of items on the GS survey (over 120) it is possible that satisficing approaches were so prevalent among respondents that the attempt to compare effort between true positives and false negatives was misguided.

Recall Explanation of Misclassification

Not surprisingly, students who had taken more than one CBL course were more likely to correctly classify themselves than students who had taken one CBL course. Consistent with the survey research literature (Groves et al., 2009) it would be expected that a student who had taken two courses with CBL components would have more relevant memories from which to retrieve information that he or she had participated in a service-learning course. First-year students, most of whom would have taken eight college courses, were less likely to correctly classify themselves than were seniors, juniors, and sophomores who likely would have taken between 16 and 32 courses. This finding is inconsistent with a "generic memory" explanation of recall failure (Groves et al., 2009). In other words, it does not seem that misclassification was due to students who had taken more courses than others having difficulty remembering salient details about particular CBL courses.

Unclear Terms Explanation of Misclassification

The internship findings suggest that it is possible for students to provide accurate answers to survey questions about some past community engagement behaviors. Part of the success of the internship item compared to the service-learning item on the general survey may be related to ease of recall— having to access memories of up to a few summers about a major activity compared to multiple courses for each semester enrolled at the institution. However, since more than one-third of CBL students misreported their enrollment for the current academic year on the CE survey, and about 70% of first-year students who had taken a CBL course reported that they had not done so, it seems likely that there is more at play than simple recall. As mentioned earlier, "community-based learning" rather than "service-learning" is typically used at this institution, so the GS survey item which used the term "service-learning" may not have been recognized as meaning "community-based learning" to some respondents. Although a greater percentage of students correctly reported their community-based learning course participation in the CE survey, it is possible that the limited time range of one year, or the response categories (yes/no) caused an increase in correct classification, and that language was still problematic.

Perhaps the most compelling explanation for the failure of the course items to collect accurate data is that "community-based learning" and "service-learning" are inappropriate stand-alone terms to use in surveys. It is possible that both of these terms are best characterized as higher education research jargon and are terms that are unfamiliar to many students. In other words, using these terms on a survey may be akin to asking a respondent, "Have you had dental prophylaxis in the past year?" (a technical term for having one's teeth cleaned). As survey research methodologists specify, for an item to produce valid data it must be understood to mean the same thing to all respondents, and survey questions must be constructed to adequately match the way respondents have understood their experiences and encoded their memories (Fowler, 1995; Groves et al., 2009). In contrast to the items asking about course enrollment, the internship items used unambiguous language of "internship" that matched repeated institutional language about the experience.

The use of "community-based learning" and "service-learning" may be particularly problematic because of the variation in language in the civic engagement field (Eyler & Giles, 1999; Saltmarsh, 2005). Given the lack of consistent terminology, it would not be surprising for faculty members to vary in their use of "community-based learning" or "service-learning" or to use neither descriptor. Furthermore, the courses in this study likely differed in their emphasis of the community-based learning component, particularly since a very common way for faculty to construct a service-learning course is to take an existing course and add a service-learning component (Tyron et al., 2008). For example, a course that requires students to engage with community members on a weekly basis may be more salient as a community-based learning course than a course in which part of a final project involves service-learning. Although analyses by course were limited, it is notable that students who had enrolled in courses with substantial community components (e.g., teaching adult literacy or tutoring high school students) were among those who incorrectly reported that they had not taken a community-based learning course during the academic year. Perhaps students who took more than one CBL course were more likely to correctly classify themselves because they were more likely to have been in a course in which the instructor used the term "service-learning," "community-based learning," or both.

Limitations

This study has several limitations. First, it was conducted at a single institution, and results may not be generalizable to other colleges and universities. Second, as described earlier, analyses of classification errors could only be conducted for the group of students who had taken a CBL course or participated in a funded internship according to administrative data. Third, the relatively small number of survey respondents who had taken a community-based learning course precluded some potentially meaningful comparisons, particularly analyses of individual courses. Finally, conclusions about the cause (or causes) of error in students' self reports are speculative, given the nature of these data.

Implications

For many researchers, the extent to which students misclassified their CBL course enrollment should be sobering. Compared to many of the questions on educational surveys, asking students about CBL or service-learning courses seems to require rather simple recall and less effort for retrieval because the experiences seem to be fairly distinctive (e.g., many often involve traveling to an off-campus location rather than only having course meetings in a college building). At this particular institution, nearly all students live on campus, take an average of eight courses per year, and experience relatively few course offerings that include meetings outside of the classroom or laboratory. The relative homogeneity of the academic experience at this institution should, theoretically, make CBL experiences particularly salient for students. In designing a survey item from an "armchair" perspective, it might seem reasonable that a student could successfully report whether or not he or she had taken a CBL course.

Although it is possible that these terms may work on other campuses at which "service-learning" or "community-based learning" are part of the student lexicon, they cannot be assumed to produce valid results. Researchers who conduct surveys at multiple institutions must be aware of differences in students' understanding of these educational phrases and construct items that describe courses or behaviors in ways that are understandable to all respondents. This suggests that survey items that attempt to measure service-learning should ask about particular aspects of students' behavior in courses (e.g., "Have you taken a course for academic credit in which you worked on a project with Springfield community members?"). This would be consistent with general survey construction practices recommended by survey research methodologists (Fowler, 1992, 1995; Groves et al., 2009;

Sudman & Bradburn, 1982). Surveys asking students about community engagement should be carefully examined and tested to maximize the likelihood that students can accurately respond to survey questions.

In addition, researchers should look to alternative data sources, such as administrative records, to understand students' community engagement activities. Of course, using course enrollment data requires that a campus is able to identify all of the courses that have a service-learning component, which may be possible at small institutions or institutions with sufficient course level information. However, these administrative data can only answer a subset of questions (e.g., what proportion of students are enrolled, how they compare to the student body in terms of gender or race). Without survey data, they cannot answer many of the other pressing questions for higher education researchers and institutions.

Part of the challenge of survey research is overcoming the belief held by many individuals that conducting a survey is a simple process (Clark & Schober, 1992). These results suggest that, at times, higher education survey researchers may be unable to model their approach after that of public opinion researchers. For many questions of interest within higher education, the specific language and terminology employed by different institutions can result in students at two different campuses understanding a particular survey question to mean two different things. Fundamentally, surveys need to be evaluated to ensure that items get at the relevant construct of interest, which, in some instances, may require deviation from fixed item language. At a minimum, survey item testing needs to take place at the campus level in order to ensure that survey language is understood by respondents. The burden of proof that surveys produce valid measures should be on the researcher. Since there is little evidence regarding the validity of community engagement survey measures, it is important that we not ignore Porter's (2011) admonition to engage in basic quantitative research in order to construct appropriate tools to gather data for researchers, practitioners, and policy-makers. To create good survey measures for the field, a group of service-learning and community engagement scholars need to undertake a rigorous survey project that begins by clearly defining a universe of behaviors from which survey items could be constructed and tested.

Notes

1. I intended to measure the number of days from when students received the survey invitation to when they completed it as a continuous variable. However, the variable had a trinomial distribution with 25% of respondents (75% total) completing the survey within one day of receiving the invitation, first reminder, and second reminder.
2. Both the GS survey and the CE survey were censuses rather than random sample surveys. No significance tests are reported because there is no sampling error (Cowger, 1984, 1985).
3. The GS survey asked about internships generally, whereas the database only includes college-funded public-service internships, so direct comparisons are nonsensical.

Correspondence

Correspondence concerning this article should be addressed to: Ethan A. Kolek, Director of Evaluation, Center for Community Engagement, Amherst College, P.O. Box 5000, Amherst, MA 01002-5000; ekolek@amherst.edu

References

Asiu, B. W., Antons, C. M., & Fultz, M. L. (1998). Undergraduate perceptions of survey participation: Improving response rates and validity. Paper presented at the Association for Institutional Research Annual Forum, Minneapolis, MN.

Astin, A. W, & Sax. L. J. (1998). How undergraduates are affected by service participation. *Journal of College Student Development, 39*, 251-263.

Bowman, N. A. (2011a). Examining systematic errors in predictors of college student self-reported gains. *New Directions for Institutional Research, 150*, 7-19.

Bowman, N. A. (2011b). Validity of self-reported gains at diverse institutions. *Educational Researcher, 40*, 22-24.

Bowman, N. A., & Bradenberger, J. W. (2010). Quantitative assessment of service-learning outcomes: Is self-reported change a reasonable proxy for longitudinal change? In J. Keshen, B. A. Holland, & B. E. Moeley (Eds.), *Research for what? Making engaged scholarship matter* (pp. 25-43). Charlotte, NC: Information Age Publishing.

Bowman, N. A., Brandenberger, J. W., Mick, C.S., & Smedley, C. T. (2010). Sustained immersion courses and student orientations to equality, justice and social responsibility: The role of short-term service learning. *Michigan Journal of Community Service Learning, 17*, 20-31.

Bringle, R. G. (2011). Preface. In R. G. Bringle, J. A. Hatcher, & S.G. Jones (Eds.), *International service learning: Conceptual frameworks and research* (p. ix-xvi). Sterling, Virginia: Stylus Publishing.

Campus Compact, (1999-2012). Who we are. Retrieved from: http://www.compact.org/about/history-mission-vision/

Chesbrough, R. D. (2011). College students and service: A mixed methods exploration of motivations, choices, and learning outcomes. *Journal of College Student Development, 52,* 687-705.

Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 15-48). New York: Russell Sage Foundation.

Cowger, (1984). Statistical significance tests: Scientific ritualism or scientific method? *Social Service Review, 58*, 358-372.

Cowger, C.D. (1985). Author's reply. *Social Service Review, 59*, 520-522.

Cruce, T. M., & Morre, J. V., III. (2007). First-year students' plans to volunteer: An examination of the predictors of community service participation. *Journal of College Student Development, 48*, 655-673.

Dillman, D.A., Smith, J.D., & Christian, L.M. (2009). *Internet, mail and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ.: John Wiley & Sons.

Dugan, J.P. & Komives, S. R. (2010). Influences on college students' capabilities for socially responsible leadership. *Journal of College Student Development, 51*, 525-549.

Eyler, J. & Giles, D. E., Jr. (1999). *Where's the learning in service-learning?* San Francisco: Jossey-Bass.

Foddy, W. (1993). *Constructing questions for interviews and questionnaires: Theory and practice in social research*. New York: Cambridge University Press.

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

Fowler, F. L. (1995). Improving survey questions: Design and evaluation. Thousand Oaks, CA: Sage Publications.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E. & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Hatcher, J. A., & Erasmus, M. A. (2008). Service-learning in the United States and South Africa: A comparative analysis informed by John Dewey and Julius Nyerere. *Michigan Journal of Community Service Learning, 15,* 49-61.

Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population. *Public Opinion Quarterly, 72*, 836-846.

Hurtado, S., Enberg, M. E., Ponjuan, L., & Landreman, L. (2002). Students' precollege preparation for participation in a diverse democracy. *Research in Higher Education, 43*, 163-186.

Hutchinson, S. R., & Lovell, C. D. (2004). A review of methodological characteristics of research published in key journals in higher education: Implications for graduate research training. *Research in Higher Education, 45*, 383-403.

Jans, M., & Roman, A. (2007). National response rates for surveys of college students: Institutional, regional, and design factors. Paper presented at the 62nd annual conference of the American Association for Public Opinion Research, Anaheim, CA.

Krathwohl, D. R. (1998). *Methods of educational & social science research: An integrated approach* (2nd ed.). Reading, MA: Addison-Wesley Educational Publishers.

Krosnick, J.A. (1999). Survey research. *Annual Review of Psychology*, *50, 537-567.

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

Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*, 63-82.

Malaney, G. D. (2002). Scholarship in student affairs through teaching and research. *NASPA Journal, 39*, 132-146.

Marks, H. M., & Jones, S. R. (2004). Community service in the transition: Shifts and continuities in participation from high school to college. *The Review of Higher Education, 75*, 307-339.

McCormick, A. C., & McClenney, K. (2012). Will these trees ever bear fruit? A response to the special issue on student engagement. *The Review of Higher Education, 35,* 307-333.

National Service Learning Clearinghouse. (2012). What is service-learning? ETR Associates. Retrieved from http://www.servicelearning.org/what-service-learning

National Survey of Student Engagement (n.d.). Benchmarks of effective educational practice. Indiana University Center for Postsecondary Research. Retrieved from http://nsse.iub.edu/pdf/nsse_benchmarks.pdf

Petrolia, D. R. & Bhattacharjee, S. (2009). Revisiting incentive effects: Evidence from a random sample mail survey on consumer preferences for fuel ethanol. *Public Opinion Quarterly, 73*, 537-550.

Pike, G. R. (2008). Using weighting adjustments to compensate for survey nonresponse. *Research in Higher Education, 49*, 153-171.

Porter, S. R. (2011). Do college student surveys have any validity? *The Review of Higher Education, 35*, 45-76.

Porter, S. R., Rumann, C., & Pontius, J. (2011). The validity of student engagement survey questions: Can we accurately measure academic challenge? *New Directions for Institutional Research, 150*, 87-98.

Porter, S. R. (2005), Survey research policies: An emerging issue for higher education. *New Directions for Institutional Research*, *124*, 5–15.

Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement and personality. *Research in Higher Education, 46*, 127-152.

Porter, S.R. & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44*, 389-407.

Porter, S. R., Whitcomb, M. E. and Weitzer, W. H. (2004), Multiple surveys of students and survey fatigue. New Directions for Institutional Research, *121*, 63–73.

Saltmarsh, J. (2005). The civic promise of service-learning. *Liberal Education, 91*(2), 50-55.

Sheatsley, P. B. (1983). Questionnaire construction and item wording. In P.H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of Survey Research* (pp. 195-230). Boston: Academic Press.

Singer, E., van Hoewyk, J., & Maher, M. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly, 64,* 171–188.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design*. San Francisco: Jossey-Bass.

Tanur, J. M. (1994). *Questions on questions: Inquiries into the cognitive bases of surveys*. Thousand Oaks, CA: Sage Publications.

Teisl, M. F., Roe, B., & Vayda, M. (2005). Incentive effects on response rates, data quality, and survey administration costs. *International Journal of Public Opinion Research, 18*(3), 364-373.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. New York: Cambridge University Press.

Tyron, E., Stoecker, R., Martin, A., Seblonka, K., Hilgendorf, A., & Nellis, M. (2008). The challenge of short-term service-learning. *Michigan Journal of Community Service Learning, 14*, 16-26.

Willis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.