2012 AAAS Measurement Workshop – Faculty Surveys White Paper
Ellen Iverson, Science Education Resource Center at Carleton College

As undergraduate education and teaching methods have come under increased scrutiny, survey approaches have become an important mechanism for understanding teaching practices nationwide. High attrition rates of students from STEM majors, particularly those of underrepresented groups (NRC,2011), combined with an impending workforce demand for STEM professionals (President's Council of Advisors on Science and Technology, 2012) make effective teaching in STEM fields critical to the success of our future. In this context, survey-based studies play an important role in measuring the impact of large scale professional development efforts aimed at transforming teaching methods (NRC,2012).

The ubiquitous availability of the internet to U.S. faculty, makes online survey methodology an obvious, scalable approach for studying STEM education. Surveys are the only tool that can be used to understanding the behavior of a large population (Ritter and Sue, 2007; Dillman, 2000). They can be used effectively even when the population is dispersed geographically. Further, based on a significant sample, an investigator can make robust inferences about an even larger population.

While relying too heavily on the self-report data of survey findings poses challenges, this paper describes a range of ways that surveys can be used, addresses challenges of survey administration and limits of survey data, and discusses strategies for overcoming these challenges. The paper is limited to a description of the ways in which surveys can illuminate an understanding of faculty teaching practice and does not address the broader use of surveys to understand student learning, retention, or community building.

Measuring Changes in Faculty Behavior

Surveys of faculty can be one of the most efficient ways to obtain self-reported information about teaching practice, changes in teaching practice, attitudes toward teaching, and other behaviors associated with teaching (e.g. participation in professional development, use of internet). Studies of K-12 teachers show that change in educational practice takes time and persistence. Before changes in practice can be measured, the professional development intervention needs time to mature. Participants must also have sufficient time to implement changes in their teaching. During this period, a strong plan and pressure on subjects to change is essential (Fullan, 2001). Surveys can provide a means of measuring this change over time. When operating at full-scale, programs aim to reach a critical mass after which a pedagogical approach starts to spread (Rogers, 1995). Surveys used throughout the lifecycle of a program can be a powerful tool to measure the rate of adoption for a given teaching technique and for recognizing tipping points.

Surveys can also be used to measure the fidelity of a program's learning goals to participants' perceptions. Key aspects of professional development programming for K-12 teachers include (Loucks-Horsley et al., 2003, Garet et al., 2001) 1) imparting a strong pedagogical vision, 2) developing new knowledge or skills, 3) providing opportunities to model effective teaching strategies, and 4) providing ongoing support. Surveys can be used to measure the success of the program in meeting these types of goals. For example, surveys have been used to measure the fidelity of participants' understanding of a particular pedagogic strategy (Dancy and Henderson, 2010).

To measure the success a program has in meeting other professional development goals may require a survey administered at different points in time during and following a program. For example , pre-post surveys of participants' attitudes have been used to measure shifts in teaching philosophy (McLaughlin et al., 2010) which may preclude any changes in practice (Loucks-Horsely et al., 2003). Surveys can be used to obtain self-report data from participants on the features of a professional development program that they perceive as supporting changes in their teaching. For example, surveys have been used to

measure the perceived success of different types of professional development activities being funded by single federal program. The survey in this case identified best practices in fostering collective participation, active learning, and integration of the new knowledge with participants' experiences and behavior (Garet et al., 2001).

Survey Types and their Possible Contexts

Surveys can be grouped by three primary goals (Salant and Dilman, 1994):

1) Community surveys gather data on community attitudes and expertise and can be used to identify motivations and barriers,
2) Market surveys determine the demand for or reach of a program within a target population,
3) Evaluation surveys gather data about the impact of a program or intervention.

All three survey types could include questions about attitudes, behaviors, and characteristics related to faculty teaching.

Community Surveys in STEM Education

To understand the teaching practices of a community of faculty, surveys are administered to large, representative samples of faculty (Dancy and Henderson, 2009; McMartin et al., 2008; Borrego et al., 2010). Prior to the development of a program, this type of survey can be used as a needs assessment or to obtain baseline data. For example, the On the Cutting Edge professional development program in geoscience obtained data via a survey on current teaching practices, barriers and motivations for making changes in teaching practice, resources used to learn about new teaching practices, and behaviors for sharing teaching practices (Macdonald et al, 2005). Results from this type of survey can be used to understanding gaps that a professional development program may be positioned to fill (Borrego et al., 2010). If the survey instrument is used repeatedly with the same sample population, analysis can reveal trends or shifts in responses at the scale of the community.

Market Surveys in STEM Education:

Since the 1910's, surveys have been the mainstay for market research (Douglas, 2009) with well-documented strengths and limitations. In STEM education, market surveys can be used to evaluate the penetration of a pedagogical innovation or teaching resource within a faculty community (Iverson et al., 2008). Market surveying techniques provide a less-resource intensive way to obtain some types of valuable information. Market surveys typically use a small representative sample to make inferences about the population and make use of "aided-brand awareness" type questions, such as, "Have you ever heard of the *STEM Faculty Development* workshop series for faculty?"

Evaluation Surveys in STEM Education

As project evaluations have become mandatory for education projects supported by the National Science Foundation (NSF TUES solicitation, 2010), evaluation surveys have come into widespread use as a technique for assessing the impact of faculty development programs or the adoption of new learning materials or strategies (ccliconference.org/abstracts in 2011). To obtain useful information, the evaluation survey must be designed in the context of the project's theory of change and measure impacts that are appropriate given the duration of the particular project. Program participants, stakeholders,

project personnel, curriculum authors, curriculum testers, department chairs, or recognized community leaders are all possible sampling pools for an evaluation survey.

Utilization-focused evaluation (UFE) carries this approach further by focusing on the realities of program implementation (Patton, 2008). Patton (2008, p 222) notes that UFE, "challenges decision makers, program staff, funders, and other users to engage in reality testing, that is to test whether what they believe to be true (their espoused theory of action) is what actually occurs (theory-in-use)." With this frame of reference, surveys can be used to gather data on how a program operates in terms of project management and communications. End of workshop surveys evaluate the immediate effects of a program activity on participants. Longitudinal participant surveys tap participants during and following their program experience. Findings from these surveys shape the project investigators' understanding of the persistence of the program influence over time.

Strengths and Challenges

The survey technique provides a time-efficient and convenient way to gain information about the reported knowledge, skills, and attitudes as it relates to a professional development STEM program, curriculum, or pedagogical approach. Surveys administered synchronously at the end of a workshop allow a common set of responses to be gathered from an entire group of participants. This is an efficient data collection method, reducing the time needed for both data collection and analysis over the time required interviewing a similar sized pool. Moreover, surveys administered asynchronously provide a means for the population in question to respond at their convenience over a period of time. A survey instrument that contains well designed questions (Fowler, 1995; Ritter and Sue, 2007; Dillman et al., 2008) provides the best means of reaching a large population of the STEM community.

The sample survey technique allows a researcher to survey a smaller sample size and make inferences about the larger population (Utts, 2005). For the inferences to be valid, the sample must be purposefully chosen to represent the larger population. The selection of the sample must use a probability sampling procedure to ensure that sampling error can be measured and opportunities for bias reduced. A probability-based sample identifies a sampling frame, the list of those in the target population, and then researchers use a randomized process to select from this frame. If the survey is given to a convenience sample, such as the faculty who happen to attend a professional association meeting, then it is not possible to make inferences about a larger population from the survey responses. Typically, the sample survey technique gathers quantitative data through forced-choice questions. The analysis of survey responses can range from a report that includes simple descriptive statistics, such as the percentages of responses from different types of institutions, to more sophisticated statistical analyses that capitalizes on the power of a probability-based sample. It is also possible to gather qualitative responses from open-ended questions. These can then be qualitatively analyzed for themes and the abundance of different types of responses reported.

While the survey measurement technique can reach the largest number of respondents in the most time efficient and scalable way, the design and administration of a quality survey is not trivial. Four common sources of error must be addressed in the design of a survey or discussed in the analysis of the results:

1) Sampling error: Unless the researcher has a means of surveying the entire population (as in a census), all surveys have a natural sampling error because only a sample of the population has received the survey.
2) Coverage error: Surveys should be administered to a sample that represents the population in question. A strong sampling strategy can reduce coverage error.
3) Measurement error: Surveys need strong question design to allow for interpretable responses and mitigate measurement error. Questions need to be validated to determine if the answers are obtaining the intended information.
4) Nonresponse error: Survey findings need to acknowledge any meaningful differences between respondents and those who chose not to respond to the survey.

Obtaining a Representative Sample

Like all measurement techniques, surveys rely on participant responses. To address the challenges of sampling, coverage, and measurement error, a survey should meet four requirements:

1) The sample is large enough that variations of interest cannot be attributed to random chance;
2) Everyone in the population has an equal chance of being selected for the sample;
3) Clear, unbiased, relevant questions encourage willing and accurate response; and
4) Characteristics of the respondents are similar to those who choose not to respond (Salant and Dillman, 1994).

For example, a survey could select a random sample of faculty from a professional association list. Because it is not possible to identify and survey all faculty, the findings would acknowledge this natural sampling error. To reduce coverage error, if demographics about the sample are known, such as institution type or sub-discipline, the sample could be assessed to ensure it represents the larger population of faculty. The questions would be designed and tested with think-aloud protocols to minimize any challenges in response interpretation. After conducting the survey, a random sample of the non-responders is selected and effort is made to gather responses from these faculty. The original respondent pool and the "non-responder" pool responses are compared to identify any meaningful differences, such as one group being more motivated to respond to a survey about teaching (Iverson et al., 2008).

While technology has simplified the administration of online surveys, potential respondents have become inundated with surveys. Institution research grapples with survey fatigue for the numerous surveys for which students are asked to respond (Porter and Whitcomb, 2005). Studies have focused on what aspects of survey design help garner stronger responses including the level of participant's community involvement and engagement (Groves, Singer, and Corning, 2000) or the participant's involvement in a particular topic (Ven Kohve, Wijnen, and De Wulf, 2002). Response rate remains a survey challenge as no single technique can guarantee the strong rate required.

When a sample population is not easily identified, intercept sampling, also known as pop up surveys (Ritter, Sue, 2007), provides another means of generating hypotheses about characterizations of users and identifying populations for follow up studies. The National Science Digital Library program provided hundreds of millions of dollars to support the creation of digital libraries and website collections aimed at improving faculty teaching practice in STEM fields. To allow for broad access to the resources, these

libraries were often completely open and required no registration by users, thus making the identification of a sample of users for evaluation purposes challenging. While intercept sampling may yield low response rates, a two percent response rate from a website that garners over a million unique visitors per year may be sufficient for follow on studies. Intercept sampling does not provide a probability-based sample. The relationship between the sample and the overall population is not established, thus there is no way of measuring bias or sampling error.

Determining Internal Validity

Validity refers to the strength of the inferences. To what extent did the survey measure what was intended to be measured. Cook and Campbell (1979) defined it as "the best available approximation to the truth or falsity of a given inference, proposition or conclusion."

The design of the survey must address possible threats to internal validity by considering: (Suchman, 1967)

1) History of respondents: Could the responses be explained by the history of events experienced by the participants, rather than the professional development program in question? For example, did a change in the tenure and promotion process in the same interval as the professional development program impact teaching changes?
2) Maturation of respondents: Do the results show the natural changes that all participants would experience over time, regardless of any intervention?
3) Selection: Could the results be explained by the way the participants were selected?
4) Attrition: If the survey is paired with an earlier baseline survey and participants of the population dropped out, how does that influence the findings?
5) Regression to the mean: If the survey is part of an experimental design intended to measure some type of learning gain in a pre and post-program context, regression to the mean may need to be taken into account in the analysis.

When the threats to internal validity are addressed in the survey design, the extent to which changes have been caused by the program or intervention can be more reliably determined.

There are two methods that deal with determining causation: using randomized experimental design or using theory-based evaluation. In a randomized experimental design, participants are randomly assigned to different conditions, called treatments. For example, a set of participants are randomly assigned to receive treatment (in this case faculty development) or no treatment. Surveys are administered pre and post-treatment to both groups. The treatment is then inferred to have caused the differences in response. This design is the gold-standard for identifying causal attribution because the randomized assignment minimizes any effect linked to differences in group member characteristics.

This design is rarely obtainable in studies involving large communities (Weiss, 2002), such as in the case of faculty development or education. More frequently, the characteristics of the treatment group are used to construct a comparable sample group. While this design provides a control, without the randomized assignment, it leaves open questions, including the role motivation to participate plays in causing any observed differences.

Theory-based evaluation examines the sequence of causes and effects for a professional development program and is a better fit for community-wide measurements (Weiss, 2002). A survey instrument that is part of a theory-based evaluation would examine:

1) Temporal order, for example, does the professional development intervention or curricular innovation precede the change in practice reported?
2) Covariation, are changes in the cause accompanied by changes in the effect? In other words, if a professional development program focused on addressing student's affect, are the changes faculty report in a survey aligned with changes that address affect?
3) Contiguity, is the proportion of the change reported from the survey consistent with the reach intended by the intervention? (Lipsey, 1993). If all of these conditions are met, alternatives cause for the effect should also be systematically be ruled out for a causal argument from the findings to be found plausible (Weiss, 2002; Lipsey, 1993).

Challenges of Self Report Data

Even with all of these challenges addressed, the survey technique relies upon self-reported information. Surveys which rely on self-report can measure what is conscious and easy to articulate (Wilson, 1994), such as the opinions and reflections of the respondent about their own teaching behavior. In cognitive studies where the researchers would like to make predictions about the conditions under which people are able to suppress a thought, verbal protocols (such as surveys or interviews) are a strong methodology. This type of data can also be useful for generating hypotheses but it brings into question the reliability of the data (Wilson, 1994).

Findings from surveys in physics (Henderson and Dancy, 2009) and biology (Ebert-May et al., 2011) find errors in the self-reported data. In a 2008 survey of 722 physics faculty respondents, nearly half of the faculty reported that they used at least one of the 24 research-based teaching strategies presented on the survey and nearly half reported familiarity with five or more of these strategies. However, the researchers found that with the Peer Instruction strategy, which was the most highly used strategy, respondents did not self-report behaviors consistent with the important components which characterize this strategy. While the researchers acknowledged that adoption of a teaching innovation may lead to modifications, the data suggest that many respondents have a less rigorous definition of the teaching strategy than was intended by the survey authors (Henderson and Dancy, 2009; Dancy and Henderson, 2010). The researchers believe that this data calls into question what can be assumed when faculty report using a teaching approach and whether actual classroom behavior is consistent with that approach. In a survey of 209 participants of the Summer Faculty Institute for Reforming Science Teaching program, more than 75% of participants reported frequent use of a cooperative learning activity following their training (Ebert-May et al., 2011). However, analysis of video recordings of participants' teaching using the Reformed Teacher Observation Protocol (RTOP;Sawada et al., 2002) showed frequent use of these activities in only 25% of cases. There are several reasons that the video comparison may not be fully accurate. Fully incorporating cooperative learning activities into a class may require more than the 18 months that elapsed between instruction and collection of video (Pfund et al., 2009). The video may not fully detect the student-to -student interactions central to the RTOP analysis. While additional work is needed to verify these results, the findings call into question the reliability of faculty self -reported survey data (NRC, 2012).

The biases that can influence self-report data are well documented in the behavior science and social science fields (Nederhof, 1985, Podsakoff, et al., 2003). One of the most common sources of bias is 'social desirability bias.' Respondents exhibiting this bias answer questions with the response they believe will be viewed most favorably rather than with the response that most accurately reflects their own beliefs or practices. Studies of religious practices in the U.S., (Presser and Stinson, 1998), of students attitudes toward environmental issues (Hawcroft and Milfont, 2010), and studies related to the polling results on political candidates (Keeter, 2008) have all found this type of bias. Over reporting use of a particular teaching strategy may reflect this type of bias.

Several methods have been developed to reduce the impact of social desirability bias.

1. The amount of social desirability bias in the responses can be estimated. Survey scales have been developed that include questions on situations that are socially desirable but are unlikely to actually occur, for example, "I am always willing to admit when I make a mistake"(Nederhof, 1985). The Crowne and Marlowe Scale (Crowne and Marlowe, 1964) is an example of this strategy. Responses to this set of questions can be analyzed against the 33 behaviors identified in this scale as predictors of this bias.
2. Survey design can reduce the likelihood of bias impacting the findings. For example, forced-choice responses can be constructed to include more than one socially desirable response or to include neutral items that are also socially acceptable (Nederhof, 1985). The question order and spacing between questions can also mitigate the effects of bias. Respondents are less likely to relate question items that are spaced throughout the survey to each other or to recall how they specifically responded to earlier questions. (Podsakoff, et al., 2003). The physics survey above demonstrates the use of this approach to detect bias.
3. Analysis techniques can lessen the effects on the findings. One technique involves directly measuring the presumed cause of the social desirability bias, modeling it as a latent construct, and allowing the indicators of the constructs of interest to load on this factor as well as the hypothesized constructs (Podsakoff, et al., 2003).
4. Researchers can include another measurement method in the analysis to discern the potential level bias at play (Podsakoff, et al., 2003). For example, survey findings of self-reported religious practices can be analyzed with church attendance records and observation studies (Presser and Stinson, 1998) to provide a more complete picture of the actual practices and discern the potential level of social desirability bias in the survey findings.

Mixed Methods: Survey Strategies as Part of a Data Collection Package

Surveys which can be powerful tools by themselves are more powerful when used in combination with other methods in a mixed method study. Mixed method designs use both qualitative methods and quantitative methods to study a particular question. A mixed method design can lessen the deficiencies or limitations of a single method, such as the self-report reliability of a survey. The approach also provides a more comprehensive understanding of the findings (Greene, 2010).

Good mixed method design relies on an overarching inquiry framework. The framework outlines how each method illuminates different points within the system and defines the relationship between data derived from different techniques. For example, in a mixed method study, a smaller sample of observational studies can be used in conjunction with a larger survey to understand the relationship between faculty's perceptions of their teaching practice and what takes place in their classroom.

Three examples of mixed method studies can be designed using frameworks that bring together methods in a holistic way, in a way that allows one method to provide data that complements another, or using an iterative design where one type of data informs the use of the next method. Each of these mixed method designs allows the collection of data that provides a more comprehensive understanding of the system. This is particularly valuable when evaluating complex phenomena like the educational system (Quinn Patton, 2011).

Holistic designs begin with an articulated theory of change and use different measurement techniques to test the smaller links in the change theory. The synthesis of all of these techniques provides a complex picture of the system. Complementary designs are used to clarify or enhance the findings of the other technique. Iterative designs have a more dynamic interplay between the techniques and require multiple iterations to see the larger picture (Caracelli and Greene, 1997).

The use of mixed methods designs has been successful in the health sciences, a system similar in scale and complexity to the educational system. Mixed method studies have proven effective in providing multiple perspectives and the rich, contextualized information needed to understand how an individual program impacts this complex enterprise. In recognition of its broad appeal, the Office of Behavioral and Social Sciences Research (OBSSR) of the National Institutes of Health recently defined best practice approaches in mixed-method design for health science investigations (Creswell et al., 2010). The ability to integrate both quantitative and qualitative data allows for deeper understanding of processes, why and how critical events occur, and a contextual description of the range of their effects. The benefits of this approach are well suited to understanding faculty practices as demonstrated in the examples below.

Examples 1: Survey used to generalize results – Faculty Participation in the NSDL project

A 2007 national survey of faculty digital resource use provides an example of a complementary mixed-methods design. In this approach one method informs or clarifies the next method leading to a more insightful understanding of digital library use. To inform the survey design, preliminary data were gathered from focus groups sampling STEM faculty from diverse institutions (McMartin et al., 2006; McMartin et al., 2008). Focus group studies allowed the researchers to avoid pitfalls in question language and improve the face validity of the question design. The focus groups findings indicated that faculty did not distinguish particular digital libraries and tended to use images, animation, and real world data to make their lectures more relevant to students (McMartin et al., 2006).

With this knowledge, a survey was designed to investigate barriers to the use of digital libraries. Respondents (4,678 respondents representing 119 higher education institutions) indicated that valuing educational digital resources did not always translate into use of these resources (McMartin et al., 2008). This gap between the value that faculty placed on a resource and the frequency of its use raised more questions than conclusive results. Results from the focus groups offered potential interpretations, such as the infrequent use of a purposefully considered resource or whether frequent use of a particular type of resource really makes good pedagogical sense. (McMartin et al., 2008). The synthesis of the survey and focus group findings together suggested new interpretations and identified areas for future exploration.

The questions from this 2007 survey were subsequently adapted for a study of the use of digital resources by social science faculty (Alter et al., 2012; Morgan et al., 2010). In this study questions were first refined by asking a small number of potential respondents to talk aloud as they read, interpreted, and answered the survey. This allowed the validity of the questions to be improved. The survey was then given to a random sample of the target audience stratified by institution type (1037 responses with a 32% response rate) to reduce coverage error (Alter et al., 2012). Results from the survey indicated that faculty did not place as much importance on students' quantitative anxiety as they did other factors when setting goals about quantitative literacy in their non-methods courses. To illuminate this finding, investigators conducted an interview study to understand how faculty from different disciplines or different institution types perceived the assessment of quantitative literacy (McMartin et al., 2012). The interview study revealed that students' quantitative anxiety was an overwhelming factor when introducing any quantitative activity in the classroom (McMartin et al., 2012).

This example demonstrates the importance and power of iterative designs that use deep qualitative data to both design surveys and to guide the interpretation of their results. With this mixed-method design, multiple interpretations of a particular question could be understood. This larger picture allowed for findings to be compared and validated within the context of faculty experiences. Without this iterative approach, the study would have concluded either with a strong description of the challenges of quantitative anxiety and no characterization of the scope of the problem, or with a flawed interpretation of the broad impact of quantitative anxiety across the disciplines.

Example 2: Survey used to obtain characterization of large community- On the Cutting Edge Project

The evaluation work of the On the Cutting Edge professional development program for geoscience faculty gives an example of a holistic mixed-method design. In this design, a theory of change identifies opportunities for measurement while the synthesis of the findings from the different techniques gives an integrated understanding of geoscience teaching practice. The On the Cutting Edge project aims to help geoscience faculty stay up-to-date with both geoscience research and teaching methods through face-to-face workshops, webinars, virtual events, and a growing website of resources.

In 2004 and 2009 the On the Cutting Edge program administered a national survey to faculty teaching undergraduate courses in the geoscience. Both surveys had significant response rates (39% for the 2004 survey and 48% for the 2009 survey) from a large sample (2207 responses in 2004 and 2874 responses in 2009). Findings from the surveys provide a snapshot of the teaching strategies faculty report using at the time (Macdonald et al., 2005; McLaughlin et al., 2010). While these survey findings represent a significant sample of the faculty population, the sample underrepresents faculty from two year colleges and may over-represent faculty who place higher value on their teaching as a part of their professional career. This group may be more inclined to stay informed about and improve their teaching practice.

In analyzing the 2009 survey results, statistical analysis was used to identify the relationship between Cutting Edge participation and different characterizations of teaching behavior and professional development behavior. On the Cutting Edge participants were more likely than non-participants to report changes in their teaching methods in the past two years, and specifically, they were more likely to report decreases in the amount of time lecturing and adding group work or small group activities. Participants also reported more use of in-class questioning, small group discussion and in-class exercises than non-participants (McLaughlin et al., 2010).

Interview studies were used to illuminate the relationship between reported practices and Cutting Edge participation. In interview data collected in 2004 from 54 workshop participants, 70% of Cutting Edge subjects described significant shifts in their attitude about teaching and while 80% attributed specific changes in their teaching to workshop participation (McLaughlin et al., 2010). While these data illuminate the perceived impact of the professional development program, they may well exhibit social desirability bias as the interviews were conducted by the project evaluator and a member of the project team. They may also confound the impact of the On the Cutting Edge program with the impact of other professional development experiences during the same interval and with the growing emphasis on engaged teaching practices across all of STEM education.

A second interview study compared responses of On the Cutting Edge participants to faculty who had not participated in the program. In this study, 38 faculty (22 Cutting Edge and 16 non-Cutting Edge) were asked to describe their best class, their methods for assessing learning in this class, and their perceptions of a research like cycle of educational innovation. Cutting Edge participants responded very differently from non-participants describing their preferred teaching and assessment strategies (McLaughlin et al., 2010).

The mixed-method design of the On the Cutting Edge study allowed triangulation of qualitative interview studies of small, purposeful samples with larger community surveys. Understanding the impact of a particular program on a complex human system with many different influences poses challenges. Through this approach, investigators examined different points in the program's theory of change, hypothesized how the program may be influencing participants, and examined possible changes in the overall community's reported practice. Again the multiple perspectives and integration of data gave a more comprehensive picture of this particular program system.

Example 3: Survey used to understand issues nationally for all faculty - Higher Education Research Institute (HERI) Survey of Faculty Teaching Undergraduates

The Higher Education Research Institute, which is sponsored by the University of California at Los Angeles, has been surveying faculty since 1978. The survey is administered triennially and has broad response with the 2010-2011 survey including responses from 23,824 faculty from 417 institutions (Hurtado, 2011). The question design is built on the previous instrument with HERI-affiliated researchers making suggestions based on their studies of faculty. There is an opportunity for individual campuses to include up to 20 local campus-specific questions. In addition to demographics, the themes for questions include items related to sources of stress, time spent preparing for class, and use of student-centered pedagogies.

While previous surveys focused on only full-time faculty teaching undergraduates, the 2010-2011 survey also included a set of questions pertaining to part-time faculty. In addition, the 2010-2011 survey analysis included comparisons between STEM faculty and all other faculty teaching practices. Faculty in non-STEM fields used student-centered pedagogies more frequently than STEM faculty in all areas with only three exceptions: student evaluation of each others' work, group projects, and student-selected topics for course content (Hurtado, 2011). Overall, male respondents of the survey used extensive lecture more than their female colleagues. This gender difference remained steady from previous survey analyses. In addition, both male and female STEM faculty extensively lecture more than their non-STEM counterparts. This finding suggests that even if faculty are self-reporting less lecture than in practice, there is still a significant opportunity for change.

The survey analysis of HERI provided normative estimates that were independent of any government study (Hurtado, 2011). Care was placed on the sampling strategy which only included higher education institutions, not post-secondary institutions which may be proprietary or vocational. In addition, four-year participating institutions were required to survey at least 35% of their full time undergraduate teaching faculty and universities surveyed 20% (Hurtado, 2011). The national reach of this survey provides an important view of faculty trends and changes in these issues through time. In addition, the campuses that receive HERI data are able to conduct analyses, providing additional opportunity to compare local data with a nationally normed set of data. Numerous presentations at AERA and AACU annual meetings demonstrate that this survey data is of high interest to understanding faculty in higher education.

Recommendations

It is clear that the survey measurement technique provide tremendous opportunity to learn about a large cross-section of faculty. However, understanding the nature of self-report data obtained by these studies is a central challenge. To what extent it is this data credible and reliable? Where is it likely to overreach? These issues provide a starting point for a research agenda.

This paper suggests that a mixed methods approach that synthesizes survey data with other types of data will provide the strongest information, particularly in a system as complex as the educational system. Mixed method strategies described above combine survey and qualitative data can be extended using the observational strategies that have called survey data into question. Work currently underway (Rutz et al., 2012) make use of artifacts, such as assignment prompts and student writing to further understand the experiences that faculty are reporting. Survey approaches that match student and faculty perceptions (Lopatto, 2004) provide an additional strategy for exploring this complex system.

Finally, the challenges of creating strong, reliable instruments as described above indicate the need for common, widely used, validated instruments for investigating key research questions. This will support deeper understanding of the overall STEM community and its practices and will provide the information needed to meet the learning needs of students and the workforce demands of the future.

References

Alter, G., Frey, W., Hoelter,L. McMartin, F. (2012). *Quantitative Literacy: It Starts with Faculty Development*. Presentation at the 2012 National Numeracy Network Annual Meeting. New York. Available: http://serc.carleton.edu/files/nnn/meetings/2012annual/preparing_students_ql_work.pdf

Borrego, Maura, Froyd, Jeffrey & Hall, T. Simin. (2010). Diffusion of Engineering education Innovations: A Survey of Awareness and Adoption Rates in U.S. Engineering Departments. *Journal of Engineering Education*. *99(3*), 186-207.

Caracelli, Valerie J., & Greene, Jennifer C. (1997). Crafting Mixed-Method Evaluation Designs. *New Directions for Program Evaluation, 74*, 19-30.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Creswell, John W., Klassen, Ann Caroll, Clark, Vicki L. P., & Smith, Katherine C. (2010). Best Practices for Mixed Methods Research in the Health Sciences: The Office of Behavioral and Social Sciences Research (OBSSR) of the National Institutes of Health.

Dancy, M. & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics 78(10),* 1056-1062.

Dillman, D.A. (2000). Mail and Internet surveys: the tailored design methods. Second edition. New York: John Wiley & Sons, Inc.

Ebert-May, D., Derting, T.L., Hodder, J., Momsen, J.L., Long, T.M., & Jardeleza, S.E. (2011). What we say is not what we do: Effective evaluation of faculty professional development programs. *Bioscience, 61,* 550-558.

Fowler, Floyd. (1995). Improving Survey Questions: Design and Evaluation. Thousand Oaks: Sage Publications.

Fullan, Michael G. (2001). Professional Preparation of Teachers *The New Meaning of Educational Change* (3rd ed., pp. 289-312). New York: Teachers College Press.

Garet, Michael S., Porter, Andrew C., Desimone, Laura , Birman, Beatrice F., & Yoon, Kwang Suk. (2001). What Makes Professional Development Effective? Results from a National Sample of Teachers. *American Educational Research Journal, 38*(4), 915-945.

Greene, Jennifer C., Caracelli, Valerie J., & Graham, Wendy F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-saliency Theory of Survey Participation. *Public Opinion Quarterly, 64*, 299-308.

Hawcroft, Lucy J., & Milfont, Taciano L. (2010). The use (and abuse) of the new environmental paradigm scale over the last 30 years: A meta-analysis. *Journal of Environmental Psychology, 30,* 143-158.

Heckathorn, Douglas D. (1997). Respondent-driven sampling: a new approach to the study of hidden populations. *Social problems*, *44*(2), 174-199.

Henderson, C., & Dancy, M. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physics Education Research, 3: 020102*.

Henderson, C., & Dancy, M. (2009). The Impact of Physics Education Research on the Teaching of Introductory Quantitative Physics in the United States. *Physical Review Special Topics: Physics Education Research, 5*(2).

Iverson, E., Fox, S., & Manduca, C. (2008). SERC Pop Up Survey Results. Northfield, MN: Science Education Resource Center. Available: http://serc.carleton.edu/files/serc/report_email_poll_2007.v2.doc

Keeter, Scott. (2008). Poll Power. *The Wilson Quarterly*, *32*(4), 56-62.

Lipsey, M.W. (1993). Theory as method: Small theories of treatments. *New Directions for Program Evaluation, 57*, 5-38.

Lopatto, D. (2004). Survey of Undergraduate Research Experiences (SURE): First findings. Cell Biology Education, 3, 270 – 277.

Loucks-Horsley, Susan, Love, Nancy, Stiles, Katherine E. (2003). *Designing professional development for teachers of science and mathematics* (2 ed.): Corwin Press.

Macdonald, R.H., Manduca, C.A., Mogk, D.W., & Tewksbury, B.J. (2005). Teaching Methods in Undergraduate Geoscience Courses: Results of the 2004 On the Cutting Edge Survey of US Faculty. *Journal of Geoscience Education, 53*(3), 237

McLaughlin, J. A., Iverson, E., Kirkendall, Randahl, Manduca, Cathryn, & Bruckner, Monica. (2010). On the Cutting Edge 2010 Evaluation Report. Northfield, MN: Science Education Resource Center.

McMartin, F., Iverson, E., Manduca, C, Wolf, A., & Morgan, G. (2006). *Factors Motivating Use of Digital Libraries.* Paper presented at the Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries.

McMartin, Flora, Iverson, Ellen, Wolf, Alan, Morrill, Josh, Morgan, Glenda, & Manduca, Cathryn. (2008). *The Use of Online Digital Resources and Educational Digital Libraries in Higher Education.* . Paper presented at The International Journal on Digital Libraries Special Issue on Digital Libraries and Education.

McMartin ,F., Morgan, M., Iverson, E., Morrill, J., Larsen, K. (2012). *Evaluating Faculty and Student Use of Digital Resources for Teaching and Learning.* Paper presented at the 2012 American Evaluation Association. Minneapolis.

Morgan, G., McMartin, F., Iverson, E., Mason, B., Hoelter, L., Morill, J.,Wolf, A. (2010). Learning about your NSDL Collection's Use and Users. Presented at the 2010 National Science Digital Library Meeting. Washington D.C.
Available:

https://www.ideals.illinois.edu/bitstream/handle/2142/17454/2010-%20NSDL_PathwaysV3.0.pdf?sequence=2

Nederhof, Anton J. (1985). Methods of coping with social desirability bias: a review. *European Journal of Social Psychology, 16,* 263-280.

National Research Council. (2011).Expanding Underrepresented Minority Participation: America's Science and Technology Talent at the Crossroads. Washington, DC: The National Academies Press.

National Research Council.(2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. Washington, DC: The National Academies Press.

Pfund, Christine, Miller, Sarah, Brenner, Kerry, Bruns, Peter, Chang, Amy, Ebert-May, Diane, . . . Handelsman, Jo. (2009). Summer Institute to Improve University Science Teaching. *Science, 324*(5926), 470-471.

Podsakoff, Philip M., MacKenzie, Scott B., Lee, Jeong-Yeon, & Podsakoff, Nathan P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology, 88*(5), 879-903.

Porter, Stephen R., & Whitcomb, Michael E. (2005). Non-Response In Student Surveys: The Role of Demographics, Engagement, and Personality. *Research in Higher Education, 46*(2), 127-152.

President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Washington D.C.: Executive Office of the President, President's Council of Advisors on science and Technology.

Presser, Stanley, & Stinson, Linda. (1998). Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance. *American Sociological Review, 63*(1), 137-145.

Quinn Patton, M. (1997). *Utilization-Focused Evaluation* (3rd ed.). Thousand Oaks: Sage Publications.

Quinn Patton, M. (2010). *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York, NY: Guilford Press.

Ritter, Lois A., & Sue, Valerie M. (2007). Using Online Surveys in Evaluation. *New Directions for Program Evaluation, 115*, 1-64.

Rogers, Everett M. (2003). *Diffusion of Innovations* (5 ed.). New York: Free Press.

Rutz, C., Condon, W., Iverson, E. R., Manduca, C.A., and Willett, G. (2012). Faculty development and student learning: What is the relationship? Change, May/June 2012, 42-49

Salant, Priscilla, & Dillman, Don. (1994). *How to Conduct Your Own Survey*. New York: John Wiley and Sons.

Sawada, D., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics, 102*(6), 245-252.

Seymour, Elaine, & Hewitt, Nancy. (1997). *Talking About Leaving: Why Undergraduateas Leave the Sciences*. Boulder, CO: Westview Press.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Knowledge and teaching: Foundations of the new reform., 57*, 1-22.

Suchman, E.A. (1967). The conduct of evaluative research and the evaluative research design *Evaluative research: Principles and practice in public service & social action programs* (pp. 82-90 and 91-114). New York: Russell Sage Foundation.

Utts, Jessica M., & Heckard, Robert F. (2006). *Statistical Ideas and Methods*. Belmont: Thomson Brooks/Cole.

Van Kenhove, Patrick, Wijnen, Katrien, & DeWulf, Kristof. (2002). The Influence of Topic Involvement on Mail-Survey Response Behavior. *Psychology and Marketing, 19*, 293-301.

Ward, Douglas. (2009). *A New Brand of Business: Charles Coolidge Parlin, Curtis Publishing Company and the Origins of Market Research*. Philadelphia: Temple University Press.

Weiss, C.H. (2002). What to do until the random assigner comes. In Mosteller & R. Boruch (Ed.), *Evidence matters: Randomized trials in education research* (pp. 198-224). Washington, DC: Brookings Institution.

Wilson, T.D. (1994). The proper protocol: Validity and completeness of verbal reports. . *Psychological Science, 5*(5), 249-252.