

Reconsidering Common Assessment Practice

John Weng

San Diego State University

Author Note

John Weng is a first year Masters Candidate in the Post Secondary Educational Leadership Program with a specialization in Student Affairs.

Rationale Statement

Throughout the past few decades, the emphasis of assessment has become an increasingly key focus within the practice of Student Affairs. The assessment process, as defined by Upcraft and Shuh (1996), is a cyclical process in which data is gathered and analyzed for the sole purpose of program improvement and implementation of new ideas. A key element of assessment is the attitude of continued information processing in order for betterment to occur. Given the competition of resources in higher education in recent years, however, practitioners are consistently receiving pressure from administration to prove the effectiveness of their programs. Furthermore, Student Affairs professionals are finding the need to prove that their programs are efficient and effective as a method of accountability to their academic counterparts in higher education (Doerfel, 2002). As a result, assessment has slowly become utilized as a method of program evaluation (Upcraft & Shuh, 1996). Given the need for assessment to serve a dual purpose in program evaluation as well as improvement, it is evident that the method in which assessment is carried out should be closely scrutinized and evaluated for validity, reliability, and effectiveness.

In current higher education practice, surveys are becoming a predominant method of data collection without any analysis of efficiency. With the increases of access to the internet and power of data collection tools and survey analysis software provided by companies such as Campus Labs, survey use is continually skyrocketing. With only support from one campus in 2001, Campus Labs now has over 750 member campuses with an array of Student Affairs software for assessment and reporting, demonstrating an alignment of data collection methods across the member institutions (“A Revolution”, 2014). That being said, while the technical data collection method might be similar, utilization of professional survey and assessment software does not necessarily guarantee the effectiveness of a survey. Through personal experience in

working with Campus Labs, technicians will review the survey for technical errors that might prevent a question from loading, but do not provide feedback on the content of a survey. This raises a concern for best practices in the use of survey.

While assessment and program evaluation is not always considered education research, many research best practices are applicable and useful as a frame for understanding the potential weakness of data that is produced by surveys. Particularly, issues exist with the reliability and validity of surveys that are utilized for program analysis. With the increased ease and accessibility of surveys through the aforementioned services, it is easy to become lax in the creation and design of surveys. In addition, the added pressure of accountability and program effectiveness can create a more eminent concern of consolidation and removal of programs, causing practitioners to emphasize generating a high quantity of data instead of high quality of data. Prior publications indicated that college student surveys lack validity (Porter, 2011) and that web-based surveys lack reliability in response (Perkins & Yuan, 2001). Other publications indicate that traditional issues such as nonresponse bias do not affect the usability of survey information (Peytchev, 2013). With a large amount of publications based on reviews and case studies, it becomes imperative to analyze research conducted and further evaluate the generalizability of the research. Thus, in this investigation, research revolving survey validity and reliability for the purpose of program evaluation and assessment will be reviewed. Specific areas of interest will include:

- Equivalent-form reliability of survey forms between web- and paper-based surveys
- Construct validity of surveys in the college population
- Generalizability of validity in college population
- Appropriate survey use and areas for improvement

Citation Reviews

1. Bowman, N. (2010). Can 1st-year college students accurately report their learning and development? *Educational Researcher*, 47(2), 466-496.

Synopsis

In this article, Bowman (2010) provided analysis on the accuracy of self-reported data collected from first-year college students in learning and development. The quantitative study was conducted across nineteen colleges and universities with a wide spectrum of university types. Through a cross analysis of data between self-report assessment and institutionally collected measures between two different time points, the article provides the finding that self-reported information have little to no correlation to longitudinally collected data. This finding was done through a comparison of assessments and validated measures such as the College Assessment of Academic Proficiency and Defining Issues Test 2 (Bowman, 2010). The study deconstructed elements of the assessment that were objective and subjective. The author noted that while there was some convergence of correlation in subjective measures, there was no convergence in objective measures. They attributed the convergence to be due to a shift in time rather than an actual correlation in data captured.

Relevance

As the very first paper that I read, it provided a rude awakening to my practices working in the field of assessment. Given the widespread use of self-report data as it stands as a method of program assessment and evaluation, it was very helpful to see research providing feedback that would indicate that the current practice is incorrect as well as validating the need to address the questions above. Specifically this article provides a need to critically evaluate the construct validity of the information that is being gathered in its current state. Given that there is little overlap in student's reported learning and their assessed learning, it gives me reservation and hesitation when I consider the usability of surveys in data collection.

Credibility

From a study perspective, the researcher did a good job of capture a wide variety of data that featured a spectrum of different university types. The author even managed to successfully and appropriately list limitations to their study on a both surface and deeper level. Specifically, the authors motioned that this study focused specifically on first-year college students and there was a possibility that this data could not be generalized to students who were more advanced. The authors, however, bring two points of consideration that could be pitfalls to their study. First, Bowman (2010) mentioned that the two assessment measures utilized as a longitudinal study were cognitive in nature and the subjective measures were intrapersonal and interpersonal in nature. Second, it was noted that it is possible that the two assessment measures, the DIT2 and the CAAP, were measuring different elements, creating reliability issues between two separate studies. In evaluating the authors limitation, a third greater concern rises, where the authors failed to realize the possibility that the assessment themselves lack validity. If surveys were being critiqued of their validity, it is possible that the assessments utilized, as a longitudinal measure may not be any more valid as a measure of longitudinal data. Given this concern, the results should be considered with reservation.

2. Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education, 44*(1), 1-19.

Synopsis

While no experiment was actually designed and conducted in this paper, Carini, Hayek, Kuh, and, Ouimet (2003) utilized quantitative collected by the National Survey of Student Engagement in 2001 to consider differences in web and paper surveys. The information used in the paper from the national survey featured 58,288 response sets with 37,682 completing the

paper version of the survey. Information was controlled based on different institution types, sizes, and student demographic information. From there, Carini et al. created a standard z-score based on the eight- interval scale responses of students. Through this study, the authors found that student responses that were completed on the web yielded responses that were significantly at a level of $p < 0.001$.

Relevance

The research question addressed by this article has the potential to validate the use of web-based surveys in assessing student needs. Given the length and detail of the National Survey of Student Engagement, a web-based assessment could save resources needed to coordinate paper responses. Validity of a national level assessment could also serve as a starting point for evaluating reliability of web-based surveys in other setting as well. Given the new surge of assessment within student affairs, there are not too many existing resources on paper-based surveys. By understanding the relationship of other educational related surveys in both the web and paper form, the utilization of web-based surveys can gain greater reliability as a data collection method within the field of student affairs.

Credibility

While this article provided interesting information regarding the effectiveness of web and paper based surveys, the effectiveness of the study warrants concern, especially in current higher education practices. First and foremost, the author of the article failed to acknowledge the possibility of the people who opted to fill out the survey online to fall into two distinct groups. That is, students who had access to an online form of the survey, and those who didn't. At the point in time that this data was collected, smartphones and technological integration in education was limited. Student who had access to technology might have been a distinct group and had vastly different experiences that resulted in different responses. Secondly, the author's use of a

standardized score, an average of an eight-interval Lickert scale, and interpretation of statistical significance, was flawed. The ordinal nature of the Lickert scale made the median a more effective descriptor. Even though the results presented are statistically powerful, the low effect sizes demonstrated the ineffectiveness of analysis used and act as an indicator that further research is needed for more conclusive evidence. In addition, with the tighter integration and availability of technology in the current education system, the impact of web-based surveys might also be impacted as the way we understand and interact with technology changes.

3. Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10626367>

Synopsis

Through the use of four independent studies, Kruger and Dunning (1999) attempted to understand the relationship between incompetence and inflation in self-assessments. By gathering responses from students who received credit as enrollees in psychology and human development courses, the researchers were able to compare students' ability to assess their performance and personal characteristics in juxtaposition to others based on their ability. The four experiments conducted were on the topics of humor, logical reasoning, grammar (3 and 4) respectively. In conducting the experiments, the researchers found that there was a strong negative correlation between perceived-ability and perceived test scores and actual test scores in students who performed in lower quartiles. That is, students who did worse, thought they did better. Researchers were also interested in finding the effects of showing low performing students correct responses after the experiment and its impact on perceived ability or test score. Results demonstrated that students who were low performing received an increase in their

perceived ability and scores when they were shown the correct answers.

Relevance

This article (Kruger & Dunning, 1999) addresses a fundamental element on the usability of self-reported data. Specifically, the authors looked at if one's incompetence can cause inflation in self-assessment. The authors argue that if one does not know grammar well, it would be difficult for them to assess how well they did on a grammar test since they wouldn't know when they made a mistake. That being said, their findings matched their original expectations in that those who scored poorly on a grammar test didn't realize that that was the case. The findings of this study have implications for common surveys utilized to assess student populations in student affairs. Specifically, this raises the concern of inflated responses or invalid responses on topics that students may not have a full grasp or understanding on. For example, when attempting to understand increase in leadership capacity or sense of belonging, if students cannot fully grasp what constructs satisfy successful leadership or sense of belonging, then it is possible that their responses would be inflated. This uncertainty should greatly impact the ability of surveys to be considered as conclusive measures.

Credibility

A concern that arises with this study revolves around the subjects the authors drafted. All four of the studies were conducted and completed on students who signed up for experiments as a part of a larger psychology or human development course requirement. This raises concerns as students who participated might have felt the pressure to report that they did well during the follow-up portion of the experiment. Furthermore, this study focuses specifically on intelligence as a construct of understanding incompetence's effects on self-assessment and not ability or other constructs. This raises a particular limit of this study. The study does not consider a situation when students may know absolutely nothing about a topic. On another level, it would

be interesting to consider the effectiveness of self-assessments when students lack other constructs instead of intelligence such as ability or skill.

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

Synopsis

In this excerpt from a book, authors Porter, Whitcomb, and Weitzer (2004) include two quantitative experiments that took place at a small liberal arts college to evaluate the effects of multiple surveys on student survey fatigue. In both experiments, students were randomly divided into groups and provided surveys. The first experiment involved two groups of freshman, one receiving only one survey and the other receiving two back-to-back surveys. The researchers found that the second survey that was completed by the group that was provided two surveys had a ten-percent decrease in response rate in comparison to the group who only received one survey. In the second experiment, seniors were separated into four groups. The four groups each received four, three, two, or, one surveys depending on group assignment. The group that only received one survey received the last one, the group that only received two surveys received the last two, and the group that only received three surveys received the last three. In this experiment it was found that each group had a general, but inconsistent, decrease in response rate and that each survey had a consistent decrease in survey response rate. In evaluating the survey decrease in response rate, response rates only lowered to around forty percent.

Relevance

This article provides interesting insight on the fatigue experienced by students that respond to surveys. Survey fatigue and nonresponse is a common issue in higher education. Understanding the cause of survey fatigue can improve response rate, which can then improve

the validity of data collected. This article is a nice reminder that statistical significance often does not take a large role student affairs assessment. Student affairs assessment projects often involve percentages and directions of distributions rather than actual statistical significance as Lickert-scale measures offer little information when analyzed for power.

Credibility

This study provides excellent insight into the effects of multiple surveys on survey completion rates. Published by a reputable source within the field of higher education, Wiley, the authors have clearly garnered the approval and attention of other editors and reviewers. That being said, while the work is insightful, the generalizability of the information is concerning. The authors attempt to evaluate the statistical significance of their results with little to no success. This statistical insignificance demonstrates that while survey fatigue is a prevalent problem, the primary effects are merely correlational but not necessarily significant for the purpose of statistical analysis. The sampling of the study is also concerning as the study was conducted on a selective liberal arts institution. While it is easier to consolidate assessment efforts in these smaller institutions, there is no mention on any control for subgroup populations that might have received surveys from other divisions or departmental units. Furthermore, the population at these institutions can be vastly different from larger public institutions or community colleges. These differences can also impact the findings of this study.

5. Tschepikow, W. K. (2012). Why don't our students respond? understanding declining participation in survey research among college students. *Journal of Student Affairs Research and Practice*, 49(4), 447-462.

Synopsis

Through the use of qualitative research, Tschepikow (2012) utilized data collected from 8 semi-structured interviews from a sample of 11 students selected at a large research-intensive

university. The students all had one year of experience living on campus due to the one-year live-in requirement. Students were selected from various gender identities, ethnicities, and academic backgrounds. Students were asked questions on the reasoning, impact, and willingness to participate in the surveys that administrators gave out to the institution. Through the interviews, themes that emerged from the study included students who thought of surveys as the key to creating change and those who thought of it as a necessary evil at an institution. Another theme found was that students were more likely to respond if they felt that their responses actually had a chance of creating change. Of the students who had responded, all of them indicated that they felt surveys were utilized for change. While some students enjoyed the survey process and saw it as a way of engaging with the campus, others saw it as annoying and tedious.

Relevance

This article provided interesting insight into what students perceive as the survey process. Campus administrators and leaders of institutions are increasingly engaging in assessment practices, and specifically web-based surveys for the purpose of understanding how programs are meeting learning outcomes and program outcomes. This article demonstrated that allowing the students to have insight to this reasoning allows for better engagement in the survey process. By providing insight into why some students might not respond to surveys and why they choose to fill out a survey, administrators and leaders are able to accordingly adjust and tailor their survey practices, specifically allowing students to see that their responses and voice has an impact on the campus community.

Credibility

The authors of this study did an excellent job of providing as much randomization and control as possible to conduct a study that provided insight to the students' perspective on the

survey and assessment practice at an institution. The qualitative nature of this study provides information that another survey or experiment might not have given. That being said, eight responses from a large institution are hardly representative of the entire campus's demographics. The eight students who responded were all involved with the campus on some level, which might provide different insight in comparison to the rest of the campus. Furthermore, the sampling occurred from during summer session, which meant that all of the students interviewed were enrolled during summer session. This raises additional concerns as the students who were enrolled during summer is typically a smaller sample that may or may not be representative of the overall population.

Integrative Summary

Given the lack of research on the topic of web-based survey validity in higher education assessment, in order to address the original research question, it was necessary to consolidate research from higher education as well as psychology. In addition, it was necessary to deconstruct elements of the question into specific components that could affect validity such as survey fatigue caused nonresponse bias, effects of incompetence on results, and students' ability to report learning and development. In doing so, common themes arise and create concern for the current practice of utilizing web-based survey assessment results as a form of evaluation of programs in higher education. On the topic of survey invalidity as suggested by the works of Bowman (2010), Kruger and Dunning (1999), and, Carini et al. (2003), each author found issues with utilizing surveys as a method of collecting data, either in higher education settings in general, student learning and development settings, and in situations that students are less comfortable with.

Given the concerns with the validity of results in web-based surveys, other articles that primarily focused on response-related issues with surveys help student affairs professionals

establish methods of increasing participation, in hopes to help improve some of the validity concerns associated with a survey measure. Porter et al. (2004)'s work suggested that saliency impacts the response rate of a survey that was then verified by the qualitative Tschepikow (2012). This created an understanding that students are more likely to respond to a survey given a personal buy-in factor. Furthermore, Tschepikow noted that saliency aside, students were more likely to provide feedback if they felt that their voices were going to be heard. With the mindset that multiple surveys caused survey fatigue (Porter et al., 2004), it is important to creatively establish a method of assessing less salient items from less well-known departments.

Consolidated assessment practices could address the topic of multiple survey fatigue by dramatically decreasing the number of surveys. Consolidating could also increase the consistency of data use, by encourage multiple departments to utilize applicable data in their assessment and programming efforts, which would allow students to establish an understanding that the data is being utilized. By including salient issues that the campus is addressing, the survey also becomes more attractive to the student.

Given the limitation of educational research and the lack of generalizability of the data reviewed, it is difficult to say that there is an answer on the original research question of web-based survey validity. Even though there are reservations with some of the research practice presented, it is evident that issues exist with the use of surveys within student affairs. This should create an understanding that web-based surveys, while informative and convenient, lack the power to serve as evaluation tools in deciding outcomes of programmatic evaluation.

Consolidated assessment efforts in survey use at institutions should allow for better response rate. If this is the case, assessment should still be utilized for program improvement and not evaluation until validity can be proven. Until then, further research is needed in understanding what might serve as better methods of program evaluation and ways of improving web-based

survey validity in higher education.

References

- "A Revolution in Assessment Practice." *Campus Labs*. Web. 1 Nov. 2014.
- Bowman, N. (2010). Can 1st-year college students accurately report their learning and development? *Educational Researcher*, 47(2), 466-496.
- Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education*, 44(1), 1-19.
- Doerfel, M. L., & Ruben, B. D. (2002). Developing more adaptive, innovative, and interactive organizations. *New Directions for Higher Education*, (118), 28.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing on incompetence lead to self inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10626367>
- Perkins, G. H., & Yuan, H. (2001). A comparison of web-based and paper-and-pencil library satisfaction survey results. *College & Research Libraries*, 62(4), 369-377.
- Peytchev, A. (2013). Consequences of survey nonresponse. *The ANNALS of the American Academy of Political and Social Science*, 645(188), 88-111.
doi:10.1177/0002716212461748
- Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, (121), 63-73.
- Porter, S. R. (2011). Do college student surveys have any validity? *Review of Higher Education*, 35(1), 45-76.
- Tschepikow, W. K. (2012). Why don't our students respond? understanding declining participation in survey research among college students. *Journal of Student Affairs*

Research and Practice, 49(4), 447-462.

Upcraft, M. Lee, and John H. Schuh. *Assessment in Student Affairs: A Guide for Practitioners*.

San Francisco: Jossey-Bass, 1996. Print.