

FEBRUARY 2014

NASPA Research and Policy Institute Issue Brief

FIVE THINGS



FIVE THINGS NOT To Do In Developing Surveys for Assessment in Student Affairs

NASPA
FOUNDATION

 **NASPA**
Student Affairs Administrators
in Higher Education

Rishi Sriram

FIVE THINGS ISSUE BRIEF SERIES

Supported with funding from the NASPA Foundation and published by NASPA's Research and Policy Institute, the Five Things Issue Brief Series is designed to connect leaders in the field of student affairs with academic scholarship on critical issues facing higher education. Intended to be accessible, succinct, and informative, the briefs provide NASPA members with thought-provoking perspectives and guidance from current research on supporting student success in all its forms. To offer feedback on the Five Things series or to suggest future topics, contact Brian A. Sponsler, series editor and NASPA vice president for research and policy, at bsponsler@naspa.org. Previous published briefs may be accessed at www.naspa.org/rpi

ABOUT THE AUTHOR

RISHI SRIRAM is assistant professor of higher education and student affairs, graduate program director for the Department of Educational Administration, and faculty master of Brooks Residential College at Baylor University. His research interests include student affairs practice; collaboration between academic and student affairs; and college student retention, engagement, achievement, and learning. He serves on several editorial boards and as director of research for the Texas Association of College and University Student Personnel Administrators. He holds an MEd in higher education and student affairs from Baylor University and a PhD in higher education from Azusa Pacific University.



Higher education is in an age of accountability, with accrediting agencies, government leaders, taxpayers, parents, and students demanding to see outcomes for expended resources (Hoffman & Bresciani, 2010). The costs of higher education have been increasing well beyond the rate of inflation, student loan debt is higher than ever, and students, families, employers, and policymakers are all expressing some concern that a college education does not necessarily result in appropriate employment after graduation (Blimling, 2013). For student affairs professionals, this age of accountability translates into an age of assessment: We must demonstrate that the work we do is producing the desired results.

Assessment in student affairs is “the process of collecting and analyzing information to improve the conditions of student life, student learning, or the quality and efficiency of services and programs provided for students” (Blimling, 2013, p. 5). Educators were calling for improving student services through evaluation and research as far back as 1937 in the American Council on Education’s *Student Personnel Point of View*, and student affairs professionals have always engaged in various kinds of research and assessment. Assessment is more important today than ever—it plays a central role in the reform movement in higher education that began in the 1980s (Blimling, 2013). But student affairs professionals may not possess the necessary skills to conduct assessment effectively (Blimling, 2013; Hoffman & Bresciani, 2010).

When student affairs professionals assess their work, they often employ some type of survey. The use of surveys stems from a desire to objectively measure outcomes, a demand from someone else (e.g., supervisor, accreditation committee) for

data, or the feeling that numbers can provide an aura of competence. Although surveys are effective tools for gathering information, many people don’t know how to create a survey that accurately measures what they want to know. Professionals may administer surveys that ask vague questions or that otherwise fail to capture needed knowledge. And once data are collected, researchers might not know what to do with the information besides compiling percentages of response choices. Better survey instruments will enable student affairs professionals to use the outcomes to drive decision making.

This brief offers five suggestions for avoiding common mistakes in survey design and use, and for facilitating the development of high-quality surveys that can be used to gather data for evidence-based decisions. Senior administrators and their staffs can use the brief as an introductory guide and a checklist of the fundamentals of survey development, implementation, and data interpretation.

Rather than merely proving that the work of

student affairs matters, a well-done assessment should focus on *improving* such work. Research and assessment moves the field forward by supplying information about what works for students, what does not work, and the causes of successes and failures. Just as medical doctors study and share the factors that promote health, student

affairs practitioners have an obligation to empirically demonstrate effective methods of promoting student learning. If surveys are a primary tool for gaining this kind of knowledge, they must be done correctly. This brief is a starting point for better surveys and more robust analysis of the data resulting from them.

Figure 1. Glossary of Common Terms in Survey Design and Implementation

Classical measurement model	a common model for developing scales that analyzes how responses to multiple items correlate with each other as a way of knowing how well the items measure the true score of the latent variable; also referred to as Classical Test Theory.
Conclusion validity	the extent to which conclusions based on the data are logical and appropriate.
Construct validity	the extent to which a scale behaves according to predictions.
Content validity	the extent to which a scale measures what it is supposed to measure.
Criterion-related validity	providing validity evidence by comparing a scale to an external criterion (e.g., another scale) with proven validity.
Cronbach's coefficient alpha	a common statistical method for measuring reliability by analyzing how responses to each item in a scale relate to each other.
Error	represents lack of accuracy in a scale by taking the difference between a perfectly accurate scale (1) and a scale's reliability (i.e., Cronbach's alpha), which ranges from 0 to 1.
Latent variable	an underlying psychological construct that a scale attempts to measure.
Likert response format	a common format for measurement that presents respondents with statements and asks to what degree they agree or disagree with the statements.
Measurement	the assignment of numbers to represent different degrees of a quality or property.
Psychometrics	a subspecialty directly concerned with the measurement of psychological phenomena.
Reliability	the extent to which a scale performs in consistent, predictable ways.
Scale	a measurement instrument composed of a collection of items that combine into a total score to reveal the level of a latent variable.
True score	the actual (not measurable) level of a latent variable.
Validity	the extent to which evidence and theory demonstrate that the correct latent variable is measured and the resulting conclusions are appropriate.

1 Don't Lose Sight of What You Want to Know

Every survey starts with a desire to know something. For example, we work in a residence hall and we want our students to feel a sense of community and connection. At the end of the fall semester, we measure their sense of belonging in their residence hall. This knowledge will help us make programmatic and policy decisions for the spring semester.

Sense of belonging is the *latent variable* we want to measure. Latent means underlying or hidden. Of course, we cannot take a tape measure, walk up to students, and ask them if we can measure their sense of belonging. Measuring something psychological in nature is quite different from measuring something physical. The process is messier, a condition that confronts student affairs professionals often in their assessment efforts.

The students' actual level of belonging is called the *true score*. We can never know the true score because we cannot enter into the minds of other people and measure the level of anything. But there is hope. Although we cannot directly measure sense of belonging, we can measure it indirectly and approximate the true score. To do this, we use the *classical measurement model*, also known as the classical test theory. The classical measurement model states that if we attempt to measure a latent variable multiple times, we can approximate how close we get to the true score.

2 Don't Just Ask, Measure

Survey is a general term referring to any type of questionnaire that gathers information from a sample group of individuals. Often, what we really mean when we use the term survey is a *scale*. The technical difference is that a survey merely collects information,

while a scale attempts to measure something by asking respondents to assign numbers that represent various degrees of a specific quality or property.

In the 1670s, Sir Isaac Newton attempted to measure things by making multiple observations and averaging the results for a simple representation of all the data. More than a century later, Charles Darwin observed and measured variations across species, which led to the development of formal statistical methods.

Research has shown that college students rarely accurately report on their behaviors, especially frequent, mundane activities.

Sir Francis Galton, Darwin's cousin, extended this idea of observing and measuring variation to humans. Thus, psychometrics was born—a subspecialty concerned with the measurement of psychological and social phenomena (DeVellis, 2012). Student affairs professionals venture into psychometrics when they design and administer surveys and scales.

What exactly is a scale? A scale is a measurement instrument composed of multiple questions or items. The responses to these items are combined into a total score (usually by simply adding them) that represents the level of the latent variable. In our example, we want to measure students' sense of belonging in the residence hall. Because we cannot know the actual level (true score) of their sense of belonging (latent variable), we use multiple items to indirectly measure it. Our scale takes the form of a series of statements such as these:

- I am making friends with other students in my residence hall.
- I am part of a community in my residence hall.
- I feel as though I belong in my residence hall.

- Other students in my residence hall like me.
- My residence hall feels like a home to me.

Some redundancy or overlap among items is a good thing. Note that none of these items attempts to measure student *behaviors*; for example, “How often do you talk with others in your residence hall?” Research has shown that college students rarely accurately report on their behaviors, especially frequent, mundane activities (Porter, 2011). Therefore, professionals should not try to measure specific behaviors; instead, they should focus on attitudes, perceptions, and levels of satisfaction. Each of the items in our example is an attempt to indirectly measure sense of belonging. But how do students respond to these items? This brings us to the issue of response format.

3 Don't Create Your Own Survey Format

Once we have a set of questions or items, we must choose a format for measurement. You might be tempted to create your own format for responses, but it is not necessary to reinvent the wheel—scholars have developed, tested, and published various formats that capture information efficiently. DeVellis (2012) offered an excellent summary of response formats, along with their strengths and weaknesses.

Some scales include a neutral choice . . . but respondents tend to gravitate toward this choice, so many researchers remove this option . . .

The most popular response format is the *Likert scale*, which uses declarative statements followed by response options that indicate varying degrees of agreement or disagreement with the statements (e.g., from strongly agree to strongly disagree). Some scales include a neutral choice (i.e., neither agree nor disagree), but respondents tend to gravitate toward this

choice, so many researchers remove this option and force the respondents to choose between agreement and disagreement. Unless specific research needs dictate otherwise, a six-point Likert scale with the following options works best: (1) strongly disagree, (2) moderately disagree, (3) slightly disagree, (4) slightly agree, (5) moderately agree, (6) strongly agree. A six-point scale provides enough options to capture various levels of the latent variable, forces respondents to choose between agreement and disagreement on each item, and does not have so many options that it confuses the respondent.

We now have items to measure sense of belonging and response choices for our items. But how do we know that these items actually capture sense of belonging? In other words, how do we ensure that our items are measuring what we intend to measure?

4 Don't Be Afraid of Validity

Validity pertains to whether we are measuring the correct variable rather than accidentally measuring something else and whether we are reaching the appropriate conclusions based on our findings. Our goal is to measure sense of belonging in a residence hall. If our items capture aspects of a sense of belonging, they are valid; if they do not, they lack validity.

There are lots of ways to think about validity. We will focus on four of them, each beginning with the letter “c”: (1) content validity, (2) criterion-related validity, (3) construct validity, and (4) conclusion validity (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999).

Content validity refers to the extent to which our scale measures what it is supposed to measure. How do we determine whether our items have content validity? Measurement will be more accurate if the latent variable is well defined and if the items on the scale directly link to this definition.

The first step in the process is to look at research literature on the topic, for help in defining our latent variable. For our example, we refer to an article in the *Journal of Student Affairs Research and Practice* that

Figure 2. 8 Questions to Ask Yourself in Developing a Survey

examines living-learning communities and students' sense of community and belonging. Spanierman et al. (2013) cited previous literature that defines sense of community as "a feeling that members have of belonging, a feeling that members matter to one another and to the group, and a shared faith that members' needs will be met" (McMillan & Chavis, 1986, p. 9). For our survey, we define sense of belonging as the extent to which a student connects with other students in the residence hall and feels part of the residence hall community.

Now that we have a definition, the second step is to use the literature as a guide in creating items for our scale. Essentially, we take what we learn about sense of belonging from the literature and turn it into statements with which students can agree or disagree. This is how we developed the statements in the bulleted list above.

The third step is to ask experts to review our items. How we define "expert" depends on the magnitude of our research quest. If we are creating a scale for distribution at the national level, it is worth asking national experts on sense of belonging among college students to review our items and provide feedback. But if we are going to use this scale for one residence hall, we can simply ask other student affairs professionals at our institution who have graduate degrees in a field related to higher education and who are familiar with sense of belonging to review our items. The point is not to create more work than necessary; we just want to ensure some level of external review and outside perspective.

Another way to demonstrate validity is to compare our survey with one that is already considered valid. This is known as *criterion-related validity*. To do this, we must find another scale that captures a latent variable similar to ours. Spanierman et al. (2013) used the sense of belonging scale created by Bollen and Hoyle (1990). This scale measures sense of belonging generally and does not specifically refer to a residence hall, so we will still have to create most of our own items. However, we could include one or two of Bollen and Hoyle's items in our scale and compare the responses. If the responses to our items and to the external items are similar, we have evidence of criterion-related validity. However, if students respond in drastically different ways to our items and to the Bollen-Hoyle

- 1 What do I want to measure?** Determine and define the latent variable.
- 2 What will I ask in order to measure it?** Use previous research on the topic to generate items for your scale.
- 3 How will I measure it?** A six-point Likert scale is a common, well-tested method.
- 4 How will I check for validity?** Have people who are knowledgeable in the content area review your items. Ask them about item clarity, item conciseness, and other possible items to include. Ask yourself what the resulting scores will mean and what conclusions you can and cannot draw from the scores.
- 5 How will I test the survey?** Organize a focus group to pilot your instrument and provide you with feedback on its clarity.
- 6 How will I administer the survey?** Plan ahead for how participants will receive the survey and how you can encourage them to respond.
- 7 How will I check reliability?** Reliability is easy to check with statistical software. If you are not familiar with such software, find someone on your campus who will not mind taking a few minutes to run it for you.
- 8 What will I do with my results?** Once you have determined that your scale is reliable, you can sum the responses of the items to get a total score that measures your latent variable. What will you do with this information? How will you determine that a specific score is good or bad? How will the information help you make evidenced-based decisions? To avoid misuse of your data, define in advance—for yourself and others—what your results are not saying. You can use statistical procedures to compare the scores of subgroups or to determine what other variables might help predict the results of your latent variable.

items, we should probably question the validity of our own items and try again. If we cannot find another valid scale that measures the same latent variable we are trying to measure, we might be able to use some form of external data. For example, if our latent variable directly related to academic success, we might be able to use GPA. If we can find an appropriate external criterion, this kind of validity is helpful to ensure that we are measuring what we intend to measure.

Conclusion validity is the extent to which the conclusions we draw from the results of the scale are logical.

A third form of validity is *construct validity*. Construct validity analyzes the extent to which a scale behaves the way it should. In other words, does it sort out results the way we would expect? We evaluate construct validity through statistical analyses such as exploratory factor analysis, confirmatory factor analysis, and principal components analysis. In these analyses, statistical software programs identify the responses that “hang together.” In our survey, we hope students will generally answer positively or negatively on all the items (item responses hanging together). This would show that the scale is truly measuring sense of belonging. But if students respond positively on the first three items and negatively on the last two items (item responses not hanging together), we must suspect that we are measuring two different latent variables. Statistical analyses require a large sample size (300 as a general rule of thumb); thus, determining construct validity is an advanced technique and not always practical.

The fourth type of validity is *conclusion validity*. Conclusion validity is the extent to which the conclusions we draw from the results of the scale are logical. In our example, if scores on our sense of belonging scale are generally high, we might state that students in our residence hall are happy. But we would be wrong: Happiness was not our latent variable. Claims about

happiness with our students lack conclusion validity. On the other hand, if our scores are generally low, we might conclude that students desire a greater sense of belonging in their residence hall. But we did not measure what students want, did we? On the basis of previous literature, we can say that a sense of belonging is an important part of the college experience, but we cannot make claims about what our students want or do not want, because we did not measure that variable. We measured students’ *beliefs about their current sense of belonging*, and our conclusions need to remain there. Our data tell us to what extent students feel a sense of belonging in their residence hall. To check our conclusion validity, we should ask whether our interpretations of the data are logical and appropriate.

Validity pertains to the fundamental issues of whether we measure the variable we intend to measure and whether our conclusions make sense. All four types of validity—content, criterion-related, construct, and conclusion—are important. However, trying to establish criterion-related or construct validity can overwhelm student affairs professionals, especially those who have little background in survey development or statistics. Student affairs professionals should start with content and conclusion validity. Those two types of validity are more conceptual than statistical in nature, and are the most important ones.

Once we are confident that our scale measures the correct latent variable, we can think about accuracy: How well does it measure the variable? We call this *reliability*.

5 Don’t Ignore Reliability

A scale with reliability is one that performs in consistent, predictable, and fairly accurate ways. Earlier, we established that we can never know the true score for our latent variable. How can we know the precision of our scale if we don’t know the true score? In other words, how can we know whether our scale accurately measures sense of belonging if we cannot peer into students’ souls? Reliability is a clever statistical analysis to get around this problem. First, accuracy is placed on a measurement scale from 0 (totally inaccurate) to

1 (perfectly accurate). Although we do not know the true score, we can assign the true score a measurement of 1, because 1 is perfectly accurate. Therefore, we now know that reliability varies from 0 to 1 and that 1 represents a perfectly reliable scale.

We may not know how each of our items compares with the true score, but we can calculate how each item relates with the other items. This is internal consistency reliability, commonly measured by a method called *Cronbach's coefficient alpha*. Cronbach's alpha is easy to calculate using statistical software that compares the relationship of the responses to each item with the responses to all the other items on the scale. After calculating all the relationships, it provides an overall reliability score between 0 and 1. Because we know that the true score is 1, the difference between our Cronbach's alpha and 1 must be due to *error* (lack of accuracy in the scale). If the Cronbach's alpha of our five-item scale is 0.76, the error (the distance from perfect accuracy) is 0.24 ($1 - 0.76 = 0.24$).

What constitutes acceptable reliability? After

decades of trial and error with scales, some rules of thumb have emerged. In general, a scale with a Cronbach's alpha less than 0.60 is unreliable. A scale with an alpha between 0.60 and 0.65 is undesirable. An alpha between 0.65 and 0.70 is minimally acceptable. An alpha greater than 0.70 is good, greater than 0.80 is very good, and greater than 0.90 is excellent (DeVellis, 2012).

We can also check how each individual item helps or hurts the overall reliability of the scale. Sometimes one item is particularly bad. We can remove that item and recalculate the overall reliability with the remaining items. Thus, we don't need to know ahead of time which of our items are strong and which are weak. Instead, we do the best we can to create good items from the literature, gather responses to our scale, and then conduct reliability analyses that tell us whether we need to exclude any items from the final version. Determining the reliability of our scale is especially important when it comes to efforts to improve our assessment techniques, because assessments are only as good as the instruments that underlie them.

CONCLUSION

If we didn't know any better, we might have asked the students in our residence hall one vague question ("What is your sense of belonging?") without any thought of measurement, format, validity, or reliability. However, using fundamental principles of psychological measurement, we now have a five-item scale that measures sense of belonging in a residence hall in a valid and reliable manner. We can calculate each student's sense of belonging score by simply summing the responses to the five items. If our reliability analysis told us that one of our items was hurting overall reliability, we would exclude that item and sum the other four. Once we have the total score for sense of belonging, we can use it in various statistical procedures to determine differences in sense of belonging between subgroups, variables that might relate to sense of belonging, or what might predict sense of belonging in students. We can administer the same scale at the end of the spring semester to identify any changes in

sense of belonging. The bottom line is that we've taken an important step to improve practice in support of students and the work we do to assist in their personal development and overall learning.

The purpose of this brief is to offer suggestions to help you avoid common mistakes and develop high-quality scales. Understanding how to measure latent variables, how to structure a scale, and how to ensure validity and reliability are vital aspects of survey development, and better surveys provide better information for better decisions. Graduate programs offer courses and even degrees in psychometrics—this brief is only an introduction. The additional resources provided in Figure 3 are for those who want to learn more about survey development and implementation.

The age of assessment and accountability in student affairs is here to stay. Rather than viewing assessment as a burden or an unpleasant obligation, student affairs professionals should embrace assessment for the sake

of improving the student experience. Assessment—thoughtfully planned, implemented, and used—is the key to progress in student retention, engagement, achievement, and learning. Assessment

can demonstrate what student affairs is doing well, highlight areas that need improvement, and help identify the programs, environments, and services that support positive educational outcomes.

REFERENCES

- American Council on Education. (1937). *The student personnel point of view*. Retrieved from <http://www.naspa.org/pubs/files/StudAff1937.pdf>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- Blimling, G. S. (2013). Challenges of assessment in student affairs. In J. H. Schuh (Ed.), *Selected contemporary assessment issues* (New directions for student services, No. 142, pp. 5–14). San Francisco, CA: Jossey-Bass.
- Bollen, K. A., & Hoyle, R. H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social Forces*, 69(2), 479–504.
- DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Los Angeles, CA: Sage.
- Hoffman, J. L., & Bresciani, M. J. (2010). Assessment work: Examining the prevalence and nature of learning assessment competencies and skills in student affairs job postings. *Journal of Student Affairs Research and Practice*, 47(4), 495–512.
- McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of Community Psychology*, 14(1), 6–23.
- Porter, S. R. (2011). Do college student surveys have any validity? *Review of Higher Education*, 35(1), 45–76.
- Spanierman, L. B., Soble, J. R., Mayfield, J. B., Neville, H. A., Aber, M., Khuri, L., & De La Rosa, B. (2013). Living-learning communities and students' sense of community and belonging. *Journal of Student Affairs Research and Practice*, 50(3), 308–325.

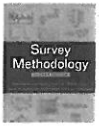
Figure 3. Additional Resources



Carpenter, S., & Stimpson, M. T. (2007). Professionalism, scholarly practice, and professional development in student affairs. *NASPA Journal*, 44(2), 265–284.



Fowler, F. J. (2009). *Survey research methods* (4th ed.). Thousand Oaks, CA: Sage.



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



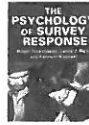
Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535.



Kane, M.T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319–342.



Smart, J. C. (2005). Attributes of exemplary research manuscripts employing quantitative analyses. *Research in Higher Education*, 46(4), 461–477.



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

ABOUT NASPA

NASPA—Student Affairs Administrators in Higher Education is the leading association for the advancement, health, and sustainability of the student affairs profession. We serve a full range of professionals who provide programs, experiences, and services that cultivate student learning and success in concert with the mission of our colleges and universities. Founded in 1919, NASPA comprises more than 13,000 members in all 50 states, 25 countries, and 8 U.S. Territories.

Through high-quality professional development, strong policy advocacy, and substantive research to inform practice, NASPA meets the diverse needs and invests in realizing the potential of all its members under the guiding principles of integrity, innovation, inclusion, and inquiry. NASPA members serve a variety of functions and roles, including the vice president and dean for student life, as well as professionals working within housing and residence life, student unions, student activities, counseling, career development, orientation, enrollment management, racial and ethnic minority support services, and retention and assessment.

ABOUT THE NASPA FOUNDATION

Higher education is being challenged on every front. Now, more than ever, it's vital that we invest in the future of our profession and the critical role we play in student success. Your NASPA Foundation advances NASPA and the student affairs profession by raising funds to celebrate achievement, support scholarships, and invest in meaningful research and emergent practices. Therefore, we are proud to partner with the NASPA Research and Policy Institute in publication of the 5 Things Issue Brief Series. This is just one example of how your gifts ensure the advancement, health, and sustainability of the student affairs profession and help answer the challenges we face as educators.

Founded in 1973, the NASPA Foundation is a philanthropic organization—our impact depends on private donations. More donors making more gifts directly translate into more research and scholarship funding for the profession. More information on the NASPA Foundation is available online at www.naspa.org/foundation, or by contacting Amy Sajko, Foundation development officer, at asajko@naspa.org.



NASPA

NASPA—Student Affairs Administrators in Higher Education
111 K Street, NE, 10th Floor | Washington, DC 20002
tel 202.265.7500 | fax 202.898.5737
www.naspa.org