

## **The effects of response set bias and international student participation on student evaluation of teaching effectiveness surveys (0272)**

laeger Paula <sup>1</sup>, Robert G. Insley<sup>1</sup>, <sup>1</sup>*University of North Texas, Denton, Texas, United States*

The use of student evaluation of teaching effectiveness (SETE) surveys is common on many college campuses. There are schools that reward faculty financially based on their SETE scores, legislative bodies that recommend a portion of faculty salary be based on such scores, faculty who cite the scores on their curriculum vitae, and colleges that post SETE scores on open websites for parents and students to review. The university in this study spent 18 months discussing SETE issues with students, administrators, and faculty through focus groups, interviews, surveys, and testing the instrument. These issues were discussed with the purpose of developing a SETE to be used across colleges at the university, regardless of the instructional delivery modes. The authors analyzed the impact of a response set bias referred to as the halo effect on the responses to the SETE and the lack of participation in the SETE by international students.

Response set bias is the systematic tendency to respond to a range of anchored survey items on some basis other than the specific content which naturally interferes with the interpretation of scores gained, when in fact they are not measuring what is intended. Haladyna and Hess (2000), Marsh and Hocevar (1991), Theall and Franklin (2001), Cooper (1981), and Feeley (2002) have all written about this bias as a halo effect and its presence in survey data collected on student evaluation of teaching effectiveness forms. Some authors like Kerlinger (1973) claim halo effect has been overrated, while Haladyna and Hess see it as a serious attack that has plagued this component of the teaching evaluation process for years. What to do with tainted responses varies from discarding them to simply noting the fact that they are assumed to be present to the degree determined through the root mean square

calculation. In response to this condition, the authors developed the Jaeger and Insley Index that analyzes specific SETE patterns and provides a mechanism to identify degrees to which the bias might be in play within a given SETE dataset.

The Jaeger and Insley Index quantifies the degree of halo effect by course and dataset, helping to reduce the impression that unless there is variability in the scores, halo effect is the cause. By pattern analysis and inter-rater scores with scaled confidence levels, it is possible to identify the degree to which halo effect is impacting the overall score of the instructor. Using the SETE dataset, the second area discussed is the bias created by the lack of participation by international students on the campus and the confusion about specific aspects of SETE that were difficult to translate causing another type of bias that Cooper (1981) referred to as lack of concreteness. Insufficient statement categories concreteness occurs when statements categories are unclear in their descriptions, making the covariance of ratings inflated due to a lack of specificity regarding rating categories. Along similar lines, the lack of concreteness can be exhibited in unclear statements, insufficient rater time expended, and insufficient rater motivation and effort.

This research is based on the data from three administrations of the new SETE (80,450), multiple survey results about issues surrounding the SETE (n=550) and data from 32 focus groups (n= 247) with both faculty and students. The focus groups helped the researchers understand the ways the students saw the use of the SETE and their use of the instrument. The range of comments about the halo effect ranged from regret for engaging in the practice and justifying it as human nature to seeing no harm in using one dominant trait to raise or lower the rating on an unrelated criterion. The impact of the halo effect will affect the usefulness of any instrument used to measure teaching effectiveness and certainly may impact the opinion of faculty members toward SETEs.

Out of the same research focus on response set bias it was discovered that when the instrument was administered by paper and pencil, fewer-than-expected international students submitted completed forms. This lack of voice or a bias of lack of participation was also studied. As a result, research translations of the SETE were made in Creole, Spanish, Chinese, Japanese, Arabic, and Vietnamese. The field tests of the translations produced strong evidence that the potential for misunderstanding the statements were high, interjecting additional bias. Through the translation project each translator spent about three hours discussing the nuances of the statements and the cultural misconceptions. Among some countries a statement that the instructor arouses my curiosity is actually a negative concept (Vietnamese) or the tone of the instrument was wrong because of the use of personal pronouns (Arabic). There are about 4,000 international students on the campus, and it is important to listen to their feedback about the effectiveness of teaching to help promote faculty development programs that address their concerns, view how the experience in the classroom is different for international students, and encourage international students to share their ideas in a classroom environment that may be considerably different than their previous experiences.

This presentation advocates for a well-designed SETE that begins with an instrument that produces more than simply means and standard deviations. The instrument at the center of this research uses scale scores that within the algorithms allow for ant colony optimization, corrections for more than 17 criteria, and data collection about the student, the course and the instructor. As the use of SETEs evolve into tools that are published on the web and used to award performance bonuses, it is necessary to analyze these instruments and conduct validity studies. From full professors to adjuncts, careers can be stifled by poorly-designed SETEs, making evidence of the halo effect a likely scapegoat for devaluing the student evaluation process. Without an understanding of the impact of the halo effect bias

or lack-of-participation bias, it is possible that poor decisions by administrators will be made as use of SETEs expands.

## References

- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218-244.
- Feeley, Thomas H. (2002). Evidence of halo effect in student evaluations of communication instruction. *Communication Education*, 5, 225-236.
- Haladyna, T. & Hess, R. (2000). An Evaluation of Conjunctive and Compensatory Standard-Setting Strategies for Test Decisions. *EDUCATIONAL ASSESSMENT*. 6, 129-153.
- Kerlinger, F. N. (1973). *Foundations of behavioral research: 2d ed.* New York: Holt, Rinehart and Winston.
- Marsh, Herbert W. & Hocevar, Dennis. (1991). Students' evaluation of teaching effectiveness: The stability of mean rating of the same teachers over a 13-year period. *Teaching & Teaching Education*, 7, 303-314.
- Theall, M. & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instructors? In M. Theall, P.C Abrami, & L.A. Mets (Eds.), *The student ratings debate: Are they valid? How can we best use them?* (New Directions for Institutional Research, No. 109, 45-56. San Francisco: Jossey-Bass.