

Margin-of-Error Calculator for Interpreting Student and Course Evaluation Data

Kenneth Royal, PhD¹

¹North Carolina State University

Overview

An online calculator was created to help college faculty and K-12 teachers discern the adequacy of a sample size and/or response rate when interpreting student evaluation of teaching (SETs) results. The online calculator can be accessed here: <http://go.ncsu.edu/cvm-moe-calculator>.

About the calculator

One of the most common questions consumers of course and instructor evaluations (also known as “Student Evaluations of Teaching”) ask pertains to the adequacy of a sample size and response rate. Arbitrary guidelines (e.g., 50%, 70%, etc.) that guide most interpretive frameworks are misleading and not based on empirical science. In truth, the sample size necessary to discern statistically stable measures depends on a number of factors, not the least of which is the degree to which scores deviate on average (standard deviation). As a general rule, scores that vary less (e.g., smaller standard deviations) will require a smaller sample size (and lower response rate) than scores that vary more (e.g., larger standard deviations). Traditional MOE formulas do not account for this detail, thus this MOE calculator is unique in that it computes a MOE with score variation taken into consideration. Other details about the formula also differ from traditional MOE computations (e.g., use of a *t*-statistic as opposed to a *z*-statistic, etc.) to make the formula more robust for educational scenarios in which smaller samples often are the norm. This MOE calculator is intended to help consumers of course and instructor evaluations make more informed decisions about the statistical stability of a score. It is important to clarify that the MOE calculator can only speak to issues relating to sampling quality; it cannot speak to other types of errors (e.g., measurement error stemming from instrument quality, etc.) or biases (e.g., non-response bias, etc.).

Persons interested in learning more about the MOE formula, or researchers reporting MOE estimates using the calculator should read/cite the following papers:

James, D. E., Schraw, G., & Kuch, F. (2015). Using the sampling margin of error to assess the interpretative validity of student evaluations of teaching. *Assessment & Evaluation in Higher Education*, 40(8), 1123-41. doi:10.1080/02602938.2014.972338.

Royal, K. D. (2016). A guide for assessing the interpretive validity of student evaluations of teaching in medical schools. *Medical Science Educator*, 26(4), 711-717. doi:10.1007/s40670-016-0325-9.

Royal, K. D. (2017). A guide for making valid interpretations of student evaluations of teaching (SET) results. *Journal of Veterinary Medical Education*, 44(2), 316-322. Doi: 10.3138/jvme.1215-201r.

Interpretation guide for course and instructor evaluation results

Suppose a course consists of 100 students (population size), but only 35 (sample size) students complete the course (or instructor) evaluation, resulting in a 35% response rate. The mean rating for the evaluation item “Overall quality of course” was 3.0 with a standard deviation (SD) of 0.5. Upon entering the relevant values into the Margin Of Error (MOE) calculator, we see this would result in a MOE of 0.1385 when alpha is set to .05 (95% confidence level).

In order to use this information, we need to do two things: First, include the MOE value as a \pm value in relation to the mean. Using the example above, we can say with 95% confidence that the mean of 3.5 could be as low as 2.8615 or as high as 3.1385 for the item “Overall quality of course”.

Next, in order to understand the MOE percentage, we must first identify the length of the rating scale and its relation to the MOE size. For example, if using a 4-point scale we would use an inclusive range of 1-4, where the actual length of the scale is 3 units (e.g., distance from 1 to 2; 2 to 3; and 3 to 4). So, a 3% MOE would equate to 0.09 (e.g., 3 category units \times 3.00% = 0.09). Similarly, a 5-point scale would use an inclusive range of 1-5, where the actual length of the scale is 4 units. In this case, a 3% MOE would equate to 0.12 (e.g., 4 category units \times 3.00% = 0.12).

Finally, we would refer to the interpretation guide (below) to make an inference about the interpretative validity of the score. In the above example the MOE for the item “Overall quality of course” was 0.1385. If we are using a 4-point scale, this value falls between 0.09 to 0.15, which corresponds to 3 to 5% of the scale (this is good!). So, we could infer the 35 students who completed the evaluation (sample) is a sufficient sample size from a course consisting of 100 students (population) to yield a statistically stable result for the item “Overall quality of course”, as the margin of error falls between \pm 3-5%. *Note:* It is important to keep in mind that 35 students are adequate in this specific example because the scores deviated on average (standard deviation) by 0.5. If the standard deviation for the item was, say, 1.0, then 35 students would have yielded a MOE of 0.2769. This value would greatly exceed 0.15, indicating the MOE is larger than 5%, and would call into question the statistical stability of the score in this scenario.

For a 4-point rating scale:

*Please note the interpretation guide does not consist of rigid rules, but merely reasonable recommendations.

Margin of Error	Margin of Error (%)	Interpretive Validity*
Less than 0.09	Less than \pm 3%	Excellent interpretive validity
Between 0.09-0.15	Between \pm 3-5%	Good interpretative validity
Greater than 0.15	Greater than \pm 5%	Questionable interpretative validity; values should be interpreted with caution

For a 5-point rating scale:

*Please note the interpretation guide does not consist of rigid rules, but merely reasonable recommendations.

Margin of Error	Margin of Error (%)	Interpretive Validity*
Less than 0.12	Less than \pm 3%	Excellent interpretive validity
Between 0.12-0.20	Between \pm 3-5%	Good interpretative validity
Greater than 0.20	Greater than \pm 5%	Questionable interpretative validity; values should be interpreted with caution

Example at NC State University:

This document [Course/Instructor Evaluation Results Interpretation Guide](#) provides a look-up chart for interpreting course and instructor evaluation results for DVM program cohorts in the NC State College of Veterinary Medicine.