Testimony of Susan Ramlo, PhD

Before the Senate Workforce and Higher Education Committee

Representative Jerry C. Cirino, Chair

April 13, 2023

Chair Cirino and Members of the Workforce and Higher Education Committee:

My name is Susan Ramlo, and I am a Professor of Mechanical Engineering at The University of Akron (UA). I do not represent UA within this testimony but, instead, my profession and expertise in higher education research.

Within this letter, I will address the well-known issues with Student Evaluation of Teaching (SET) instruments – regardless of the statements within the SET. Within SB 83, Section 3345.451 (A, B, C, D, & E), it is suggested that SET are important for evaluating the teaching quality of faculty in Ohio's public institutions of higher education and should be instituted state-wide with publication of numerical values for each faculty-person annually.

Sec. 3345.451. (A) As used in this section, "state institution of higher education" has the same meaning as in section 3345.011 of the Revised Code.

(B) The chancellor of higher education shall develop a minimum set of standard questions for use by state institutions of higher education in student evaluations of faculty members. The questions shall include the following: "Does the faculty member create a classroom atmosphere free of political, racial, gender, and religious bias?"

(C) Each state institution of higher education shall establish a written system of faculty evaluations completed by students with a focus on teaching effectiveness and student learning. Each state institution shall include in its student evaluations of faculty the minimum set of standard questions developed by the department in division (B) of this section.

(D) Not later than August 1, 2024, the average annual numerical score from the student evaluations for each faculty member shall be published on a public portal on each state institution of higher education's web site. The scores shall be updated by the first day of August of each year thereafter.

(E) Each state institution of higher education shall establish a written system of peer evaluations for faculty members with emphasis placed on the faculty member's professional development regarding the faculty member's teaching responsibilities.

Certainly, SETs are commonly used to evaluate faculty at most institutions of higher education in the US. However, their common use is not an indication that these are valid or reliable measures of faculty performance in the classroom. SETs have been used for nearly a century yet the problems with SETs have been known for decades and there is considerable research about them as inappropriate for measuring or even hinting at quality of teaching. Instead, *regardless of the items* within an SET, these instruments are typically indicators of

student biases as well as their perceived satisfaction based upon the grade they expect to receive or the easiness of the course.

Validity and Reliability

Within social science research, any instrument that purports to measure a certain characteristic or quality, must be shown to be valid and reliable. Validity is an evaluation that indicates that the measurement tool (e.g., SET survey) measures what it says it measures. Thus, if an instrument states that it measures student motivation, validity evaluations (of which there are several types) would be carried out – often mathematically. Reliability is a separate evaluation of an instrument. Reliability is the consistency of the measurement. Whereas a valid instrument is automatically deemed reliable, a reliable instrument is not necessarily valid. The evaluation of instruments, such as SET, is called psychometrics. SET are very rarely evaluated for validity and reliability. When such psychometric evaluations have taken place, regardless of the items within an SET, these evaluations have shown that SETs are poor measures of teacher quality. Instead, other (non-quality of teaching) characteristics of the faculty-person and/or the student are measured by SET. Some have found that SETs are aligned with student satisfaction, including their perception of how easy the class is perceived. SET are also an indicator of student biases.

Inherent Biases in Student Evaluation of Teaching

Faculty are often "punished" by students if they possess high student expectations and expect students to put forth considerable effort. SET has an adverse impact on female faculty as well as faculty of color. These findings are especially true for those female and faculty of color who teach within disciplines that are perceived as predominantly male or white-male. Research has also shown that instructors may attempt to receive better SET scores by becoming lenient graders and/or reducing the quality of their teaching. I don't think that this is the outcome anyone wants but it is the reality, especially when SET becomes a high-stakes evaluation tool.

Statistical Improprieties

Within statistics, the choices of the types of statistical analyses to perform are based, in part, on the type of data. Continuous data can take any value, including fractional parts. Weight and temperature are good examples of continuous data. Ordinal data is not a type of continuous data. Instead, ordinal data uses variables to represent into descriptive categories (e.g,, 5 for most agree, 1 to represent most disagree). Student Evaluation of Teaching instruments collect ordinal data for a number of items (statements) with scales of 1 to 5 or 1 to 7.

In SET, averages (mean values) for each item are typically calculated and reported as well as a singular average value for all items across all sets of data. However, statistically, averages (mean values) are only appropriate for continuous data. Thus, calculating the mean values for ordinal measures represents statistical impropriety.

Additionally, results of SET are unlikely to be gaussian – meaning a normal distribution about a singular peak value. Instead, if one plots the ordinal values from SET (whether for an individual item or for the entire instrument), the results are more typically bimodal (two peaks) or trimodal (three peaks). This type of distribution is an indication that more than one view

exists about the instructor. A gaussian distribution is another requirement for using the mean for continuous values. This finding also represents the reality that different students within a classroom, experiencing the same instructor, assessments, etc., have different subjective viewpoints about that experience.

Peer Evaluations

Peer evaluations are often considered better than student evaluations of teaching. However, biases are still present within these types of evaluations of quality of teaching. Some of these biases are the same as those seen from SET including biases against women and faculty of color. Additionally, personal subjective viewpoints about teaching best practices, that are not based on research but, instead, personal biases and preferences, also affect the outcomes of peer evaluations.

Closing

Ohio's faculty play a key role in the educating of our students – ensuring that they are career-ready and prepared to be good citizens within our democracy. While quality of teaching is a key component of our profession, student evaluations of teaching (SET) and peer evaluations are imperfect, to say the least, in evaluating the quality of the teaching we provide students. You may wonder why institutions continue to retain SETs as a measure of quality teaching so I offer what an upper administrator at my university once told me, in response to my critique of SET, "Sure, it might be wrong, but it is easy." The proposal that our teaching can be boiled down to a singular numerical value is both an insult to the students we serve as well as a means of further implementation of inherent biases that are already present in relation to the evaluation of faculty. Certainly, dependence upon SET and peer evaluations will decrease the already problematic representation within the faculty of people of color and women. In turn, this will impact Ohio institutions' ability to attract qualified faculty to educate our students and provide important research.

Thank you for your time. I welcome any questions.