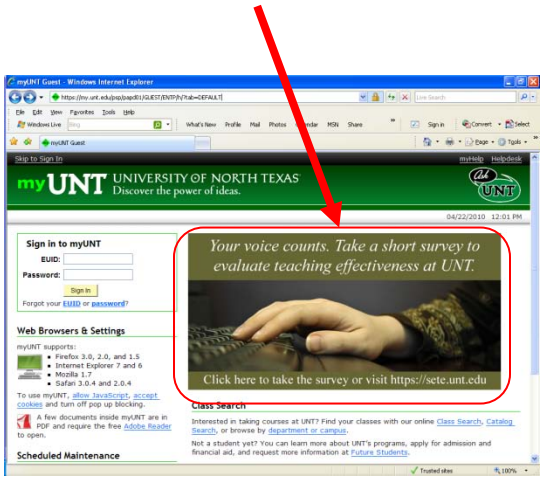
UNIVERSITY OF NORTH TEXAS
Institutional Research & Effectiveness

Spring 2010 Student Evaluation of Teaching Effectiveness (SETE)

Where do my students log in to complete the SETE?

Students can access SETE here: https://sete.unt.edu or by clicking on the SETE banner at https://my.unt.edu



Where can I see how many of my students have completed the SETE for my class?

Instructors can use their EUID and password to see the number of completions here: https://sete.unt.edu/count

What are the administration dates?

The Spring 2010 administration period is from April 19th, 2010 to Friday, May 14th, 2010.

What is the purpose of the SETE?

The original SETE committee was charged with providing to the Provost of the University of North Texas (UNT) a recommendation for an assessment tool to facilitate student evaluations

of their instructors, allowing university-wide comparison in key areas. The SETE purpose is provide a measure of teaching effectiveness as perceived by students. The SETE scores for a particular instructor can be used for self evaluation and improvement and for measuring improvement over time. The scale scores can also be aggregated into group scores for use by administrators. In addition to providing needed information for UNT, the SETE also satisfies the requirements of House Bill 2504 that calls for transparence in reporting and posting to the web.

How were the SETE items selected?

After a review of the literature and input from committee members, it was determined by the committee that the survey should focus on measuring teaching effectiveness and that course effectiveness should be treated separately and by a different committee. It was also determined that the survey instrument should be structured on the dimensions and elements presented on pages 51-53 in Berk (2006) as synthesized by Davis (1993) from research on good teaching (Chickering & Gamson, 1991; Eble, 1988; Murray, 1991; Reynolds, 1992; Schon, 1987) and on student achievement and success (Noel, Levitz, Saluri, & Associates, 1985; Pascarella & Terenizini, 1991, Tinto 1987). Berk's book, *Thirteen Strategies to Measure College Teaching,* was selected as a handbook and guide for the project.

It was determined that every effort should be made to find extant surveys and published lists of survey items and to evaluate them for usefulness versus writing new items. Major pieces of validity evidence for the results of this task are UNT faculty and student input regarding the dimensions, elements, and existing statements. This evidence was collected through seven faculty focus groups, four student focus groups, faculty and student interviews, results from a survey sent to all faculty, surveys sent to students, and an item tryout field test administered to students. Additional evidence is the scoring rubric results from the committee members' evaluation of items.

An item selection process started with a pool of approximately 3,000 survey items, including all current UNT department surveys and items published by other universities that are used by over 100 universities. After an initial screening process, this large pool was narrowed to 1,488 items. Evaluating these items with rating scales reduced this number to 788, and a second evaluation matching items to specific elements reduced the number to 346. Using specific scoring criteria to qualify items for inclusion, committee members reduced the number of items

to 51. These 51 items were then presented to students in a developmental field test and a final draft selection of 38 items was based on faculty scores and field test results. A final review was conducted using the criteria of student viewpoint, student observable, statement measurability, conformity to the research elements, duplicity, and universality in terms of class size and in terms of online and in-class administration. The result of this process is the survey final item pool of 28 statements. Over 400 people were involved in the process.

How were the final survey items chosen?

The second phase of the SETE development included three teams made up of faculty and staff who specialized in assessment development and psychometrics. Team A conducted the psychometrics; Team B administered spring pilot tests of the SETE items and conducted follow-up faculty and student focus group; and Team C developed open-ended response items. The final 28 SETE items were pilot tested using a stratified sampling across the University. The pilot test was administered at the end of the Spring semester 2009, and a validity study team was assembled to analyze the data, validate the model fit, conduct item reduction studies, and develop a scoring methodology. The result of the psychometric work was the 12 item survey that was administered across the university in the fall of 2009.

Why was a four point scale used for the SETE?

Research shows that after five points there are diminishing returns in terms of reliability. Additionally, information may be lost if the scale exceeds the respondents ability to discriminate among the anchor points. A 28 item survey with a 4-point scale can yield high reliability coefficients. It was determined that four anchor points were appropriate using a response scale of 1) Strongly Disagree, 2) Disagree, 3) Agree, and 4) Strongly agree.

Why is there no midpoint position on the scale (i.e. neutral, uncertain, or undecided)?

Information is lost when a midpoint position is included in a set of bi-polar (i.e. both positive and negative) anchors that are intended to measure the degree (intensity) of a respondent's opinion. The neutral mid-point is also problematic because it will lower the mean for a teacher who receives a high score and adds no compensation for a teacher who received a low score. From a measurement viewpoint, nothing is gained from a neutral response. Berk (2006) states

that, "For rating scales used to measure teaching effectiveness, it is recommended that the *midpoint position be omitted* and an even-numbered scale be used, such as 4 or 6 points."

Why is there no NA (not applicable) choice?

The use of NA was avoided because the teacher effectiveness scale will be used for a class level analysis, and every time a student chooses NA, that student's scale score will be different because one or more of the items will not be part of the score. This is a major problem in terms of measurement, analysis, and validity. Recognizing that there are class conditions across the university (even on the teacher effectiveness only scale) that would require an NA option, the committee followed recommended procedures for identifying which items might require an NA so they could be eliminated from the final item selection. These procedures included faculty and student review groups in which Faculty were asked to identify those items which they felt could not be observed by students across all classes and thus would require an NA, and students were asked to identify those items which they felt could not be observed by students across all classes and thus would require an NA. Identified items were eliminated.

Why is there no item reversal (negative and positive items) to address response set bias?

This type of bias is referred to as acquiescence, the tendency to agree or give positive responses regardless of the content of the items (similar to Halo effect). A strategy used to minimize the effect of this survey taking behavior is to word half of the statements positively and the other half negatively (but in random order). However, this method does not eliminate (or reduce) the bias, it simply cancels out the effect of the bias with the result that the effect of the bias is reduced to zero. Berk (2006) recommends that reversals may be appropriate for some scales, but not for teacher effectiveness scales because the positive/negative reversals can be confusing and result in increased response time and response errors. The SETE effectiveness scale is designed to rate the teacher's positive behaviors, not negative ones.

What is the applicability of SETE items to courses delivered online?

Application of the SETE items to online courses was a major consideration of the committee. Expertise in delivering online instruction was well represented in the committee. Additionally, input was gathered from faculty and student groups. Several online courses were included in the SETE field test in order to do a comparison of online versus not-online student responses. The structural equation modeling being used to confirm the structure of the student responses will include the online courses. Faculty and student review groups were convened at the beginning of the fall semester 09 to confirm final recommendations regarding the usefulness of SETE survey items for online courses.

What are the factors or elements of teaching effectiveness that are being measured?

In addition to the overall construct of teaching effectiveness, there are three specific factors or dimensions that are being measured. They are shown below along with the items that are used to get a measure of the factor. The four items used for each of the factors are the four that best fit the bi-factor structural equation model used in the SETE validity analysis. The twelve items on the SETE were chosen from the final pool of 28 usable items and the four items chosen for each factor best represent the factor. For your convenience the SETE factors with their items are provided here.

Factor 1: Organization and explanation of materials.

1. My instructor explains difficult material clearly.
2. My instructor communicates at a level that I can understand.
3. My Instructor makes requirements clear.
4. My instructor identifies relationships between and among topics

Factor 2: Learning Environment

1. My instructor establishes a climate of respect.
2. My instructor is available to me on matters pertaining to the course.

3. My instructor respects diverse talents

4. My instructor creates an atmosphere in which ideas can be exchanged freely.

Factor 3: Self Regulated Learning

1. My instructor gives assignments that are stimulating to me.

2. My instructor encourages me to develop new viewpoints.

3. My instructor arouses my curiosity.

4. My instructor stimulates my creativity.

What do my SETE scores mean? How should they be interpreted?

Your SETE scores are a measure of your students' perception of your teaching effectiveness. The scores are based on a scale across the University. In other words, all individual scores are on the same scale so that a score of, for example 600, for a teacher of a particular course in a particular department or college has the same meaning in terms of teaching effectiveness as a teacher of a particular course in a different department or college. To help with score interpretation, the following factor descriptions of effectiveness are provided on the individual teacher reports.

Factor1: **Organization and Explanation of Materials**

This score reflects the student's perception of how well the instructor: makes the course requirements and student learning outcomes clear to the students; gives assignments, activities, and materials that are helpful and that contribute to understanding the subject; explains difficult material clearly; shows the relationships among topics and new concepts; and evaluates student work in ways that are helpful to learning.

Factor 2: **Learning Environment**

This score reflects the student's perception of how well the instructor: establishes a climate of mutual respect and encouragement; motivates students to work and engage in learning; is available and encouraging; is skillful in actively engaging students in learning; and provides useful feedback.

Factor 3: **Self-regulated Learning**

This score reflects the student's perception of how well the instructor guides and encourages self-directed learning in which the student is encouraged: to be open to the viewpoints of others; to develop new viewpoints; to connect course topics to a wider understanding of the subject; and to contribute to the learning process.

Each of the three effectiveness factors has its own unique scale and thus each teacher gets a separate scale score for each factor. The overall construct of Teacher Effectiveness also has its own scale score, and thus is not simply the average of the factor scores.

To give meaning to the scores in terms of a scale of teaching effectiveness, cut points were calculated so that ranges of effectiveness could be identified. Three levels were identified and the scale score range for each level was provided as follows:

SETE Scale score ranges for effectiveness levels by factors

	Organization and Explanation	**Learning Environment**	**Self-Regulated Learning**	**Overall Effectiveness**
Highly Effective	710 - 981	659 - 972	747 - 998	702 - 998
Effective	438 – 709	347 - 658	495 - 746	406 - 701
Somewhat Effective	167 - 437	35 - 346	243 - 494	111 - 405

An important feature of the SETE scale is that it is equal interval so that a ten point change from, say, 250 to 260 is the same amount of improvement in teaching effectiveness as a change from 460 to 470. Of course, this doesn't mean that the SETE scores have no degree of error (no error is not possible), but it was possible to reduce error to a minimum and can support this with some very sophisticated psychometrics.

Were my SETE scores adjusted for class size, graduate or undergraduate level, and other variables?

The scores were not "adjusted" as such, but bias related to the variables was controlled for. A measurement model with appropriate external control variables was used in determining how items should be weighted when calculating individual scale scores. This estimation process provides a reasonably fair and unbiased estimate of the individual scale scores as well as providing a high degree of reliability and generalizability to the scale scores. Put another way, on the across the campus SETE scale, the score that reports the opinion of how the students in a particular class viewed the teacher included controlling for the variables that would most bias the student's perception. Refinement of this process will continue for the subsequent SETE administrations, but it can be demonstrated that the scores for the fall 09 semester are highly valid measures of the construct of teaching effectiveness for the UNT campus. The only caution is that a score for a class in which there was a small response rate may not be interpretable in terms of the intended purpose of the SETE.

The validity team originally ran calculations on more than 100 variables such as class size, level, etc. prior to the final calculation of the factor score that was used to develop the final scale score. Ultimately the following 17 variables were found to be the best controllers and were used in the calculation of the fall 09 factor scales.

SETE External Control Variables used to control for bias

Department; Course College; Course Size; Course Time; Course Enrolled; Course Internet; Other; Room; Faculty Gender; Faculty Age; Faculty Rank; UNT Years; Faculty Status; Faculty Type; Student Credit Earned; Student Course Load; Student Gender; Student Academic Level; Mean GPA of Student; Student Anticipated Grade

How were the SETE scores calculated?

In simplest terms, factor loadings were multiplied by mean scores to produce the factor score which was then standardized on a scale with a mean of 800 and a standard deviation of 125. Here is a basic explanation that is a little more technical for those who might be interested. The SETE responses were modeled using a number of psychometric modeling methods. A latent

variable measurement error model was derived that maximized the reliability and validity of the effectiveness constructs represented by the SETE items. A model with one overall general teaching- effectiveness domain and three teaching sub-domains were modeled by the final best fit model. Specifically, structural equation modeling methods (SEM) were used to fit a hierarchical factor model where the three sub-domains and the general domain are independent of one another (a "bi-factor" confirmatory factor model). Furthermore, best practice survey sampling methodologies were used to control response bias (e.g. sample selection bias) due to the non-random sampling design employed to collect the data. Specifically, optimal matching techniques were used to balance the data with regard to what is referred to as "external control variables" (class size, time of day, dept., college, etc.). Cut points were derived from the SETE factor scores (scaled scores) to create three proficiency levels (highly efficient, efficient, somewhat efficient). Additionally, bootstrap validation studies were performed to validate the predictive accuracy of the resulting SETE scores and proficiency levels. Finally, descriptions of the teaching domains and proficiency levels were provided for qualitatively interpreting the meaning represented by the SETE scaled scores and proficiency levels.