ECS Task Force on Implementation of Universal Student Evaluations of Teaching (USETI)
Report to ECS, October 10, 2014; Revised November 14, 2014
Written by Edward Aboufadel, Professor of Mathematics, USETI Task Force Chair

In this report, I will provide a timeline of our work so far and possible directions from here.

Before January 2014: For several years, faculty groups have been discussing the use and implementation of student evaluations of teaching (SETs) at GVSU. For instance, in 2006, a CLAS task force explored questions of online and mid-semester administration of evaluations, and the adoption of a standard form throughout CLAS. During 2011-13, the FTLC-AC studied a variety of instruments that could be adopted by GVSU, as well as platforms to administer the instrument.

Members of the Task Force believe that a key factor driving these on-going discussions for several years is that the current state of affairs at Grand Valley is no longer acceptable. At this point in time, there are over two dozen evaluation forms in use on campus, and it is difficult to say that any of them have been analyzed for reliability or validity. There does not appear to be a methodology to what we are doing. In addition, there is no standard approach to the use of SET reports in faculty evaluation (annual review or personnel review), and the practice in individual units is driven by historical precedent or idiosyncratic initiative. Consequently, when it comes to the question of adopting a standard form to be used throughout the University, we must also ask how to modernize our use and understanding of these forms.

January 2014: The following motion was passed by UAS:

The university should adopt a standardized measure of student evaluations of faculty teaching that meets contemporary standards for reliability and validity in psychometric measurement. The measure should yield both quantitative and qualitative results. The university should also adopt a standardized platform for administration of the measure (e.g., online software). The measure and platform should be used in all units and colleges. Units and colleges may employ additional instruments separately from the university standard.

March-April 2014: The Universal Student Evaluations of Teaching (USETI) task force was established, charged to make recommendations for (A) policies regarding student evaluations of teaching to be used across the university, (B) a standardized measure to be used across the university, and (C) an implementation plan. During winter 2014, task force members read some of the review research literature on SETs, and USETI determined how to proceed to meet our charges by starting with (B). A number of candidates for a universal evaluation instrument were identified (see next paragraph), and three subcommittees were created to work over the summer. The subcommittees were charged to address the following: (1) to compile evidence of instruments meeting the validity criteria: "contemporary standards for reliability and validity in psychometric measurement", (2) to compile information on other important criteria such as online delivery platforms, cost, and reporting of results, and (3) to focus on establishing an on-going dialogue with university faculty about SETs.

At the end of winter semester, based on our literature review and Internet searching, the following candidates were identified: *IDEA*[1], *SEEQ*[2], *CIEQ*[3], *IASystem*[4], *TBI*[5], *SIR II*[6], and *PICES*[7].

May-August 2014: Over the summer, USETI met twice, and at the same time, each subcommittee worked diligently on its charge, presenting a report to the rest of USETI in early August. For group #3, the report was a PowerPoint presentation that was subsequently presented to faculty at each college in August or September. A web site with information about the work of USETI was also created[8], and the presentation is available there. For group #2, information about each instrument was summarized in a PDF. For group #1, a narrative describing validity results that were found was written for each instrument.

September-October 2014: USETI has already met several times during Fall 2014 to decide on an instrument, and by mid-September, through our discussions, we have narrowed our focus to two candidates: *SEEQ* and *IASystem*. The other candidates have been rejected either because there was little or insufficient evidence that the instrument met "contemporary standards for reliability and validity in psychometric measurement", or, if there was some evidence, other factors such as cost, survey length, or lack of clear implementation at other institutions made the instrument a poor choice for Grand Valley.

SEEQ was created in the early 1970s at UCLA following the recommendations of a university task force. Like many SETs, the questions focus heavily on instructors' teaching-relevant actions, as well as aspect of course design. A large pool of items was generated based on literature reviews and interviews with students and faculty members. These items were administered to students in a large number of classes and items were retained depending upon 1) student ratings of importance, 2) faculty members' rating of usefulness, 3) factor analysis and 4) reliability analyses. Factor analyses of the SEEQ reveal nine factors: learning, enthusiasm, organization, group interaction, individual rapport, breadth, examinations, assignments, and overall evaluation. There is information available from other institutions explaining how SEEQ has been used to enhance teaching effectiveness.

IASystem was initially developed in the 1980's, and is implemented and made available by the University of Washington. Faculty, students and administrators were consulted and then potential items were critiqued relative to a specific course taught or taken by faculty and students. The IASystem is not one form but a set of forms developed for different types of classroom instruction. The IASystem does have four items that are common across all versions. The summative evaluation is based on these four overall items and the formative evaluation can be based on item sets that are available and constructed for different types of instruction or items can be selected individually by instructors.

---

[1] http://www.ideaedu.org/
[2] http://www.usask.ca/vpteaching/seeq/the-seeq-instrument.php
[3] http://www.cieq.com/images/scan_doc0025.pdf
[4] http://www.washington.edu/oea/services/course_eval/about.html
[5] http://www.calvin.edu/admin/provost/documents/behaviors.pdf
[6] http://www.ets.org/sir_ii/about/research/
[7] http://www.purdue.edu/cie/data/pices.html
[8] http://www.gvsu.edu/useti/

Attached are the validity and other criteria reports for both instruments, as well as sample forms.

<u>As of October, 2014 USETI was unable to choose between these two finalists.</u> The discussions on the task force had revolved around the extent to which there is sufficient empirical evidence that instruments meet "contemporary standards for reliability and validity in psychometric measurement" and the extent to which the questions of the instruments would be considered appropriate survey questions. When Ed Aboufadel consulted with Karen Gipson in late September, it was decided to bring the question to ECS and UAS, and to consider new members to joining USETI.

November 2014: After discussion at three different ECS meetings, at the November 14, 2014 meeting, ECS voted to re-affirm the desire to have a single instrument on campus, and to recommend that GVSU adopt *IASystem*. This recommendation will now go to UAS.

Regarding Other Recommendations: The original plan for USETI was to first select an instrument and then to develop recommendations for its implementation. At this point in time, during our discussions trying to select an instrument, or in our meetings with faculty in various colleges, a number of ideas emerged that could be part of an implementation plan. Although USETI never voted on any recommendations for an implementation plan, there are a number of ideas that we feel are worthy of further discussion by USETI going forward:

- An important point is that the implementation of a standard form should include dialogue on what SETs can and cannot tell us about the teaching performance of a faculty member – to recognize the limitations of any form that we use.
- A key question to answer is <u>how</u> results from SETs are used in faculty evaluation, both formative and summative. In the Faculty Handbook, we read, "Effective teaching must be documented by: a) self-evaluation, b) peer evaluation, and c) student evaluations." Without putting too fine a point on it, there was general agreement on USETI that something in the neighborhood of 35-40% of a summative faculty evaluation should be due to results from student evaluations.
- There was also general agreement on USETI that we don't support the idea of reducing the evaluation of teaching to a single number and using fine gradations to classify faculty as "exemplary" vs. "satisfactory".
- Another key question is to what parts of an SET report would be included in the evaluation process (hence, summative), and which would be used for the improvement of teaching (i.e., formative), accessible only to the faculty member, the Unit Head, and administrators. Some have suggested that written comments should only be formative, others ask if results for certain, specific quantitative questions make more sense as formative feedback only. There is also the case to be made that all the information in reports would be included in evaluation processes.
- Also to consider: to what extent would an SET report present unit-, college-, or university-wide data (averages or medians) for comparison purposes.
- A question of interest to many faculty members is to what extent student evaluation results are biased and not inclusive, or could potentially result in adverse impact. There

is a small set of studies on this question, but USETI should explore these issues. It is our duty that people are not mistreated by the choice and implementation of a universal SET.

- Faculty across the University are concerned about ensuring robust response rates if an online delivery system is used, and an implementation plan would need to have several solid ideas as to how to maximize response rates. We are cautiously optimistic that some of the availability of some of the online systems via smartphone and tablet applications will help improve response rates. A recent document with some good ideas is available.[9]
- A set of five-minute videos could be created to inform students about SETs – what is appropriate or not in written comments, and how the results are used by the University. These videos could be shown in class before instructing students to complete the evaluation.
- The idea of "confidential but not anonymous" evaluations could be explored, where students' names would be attached to their evaluations, in case a Unit Head or administrator needs to follow up on concerns, but the names would not be shared with the faculty being evaluated. Having students attach their names to evaluations may also encourage students to respond carefully.
- An implementation task force should re-read the work of the CLAS task force and FTLC-AC for ideas.
- A recent manuscript[10] from two Statistics faculty members at Berkeley has some other ideas to be considered in implementing SETs. (This manuscript was featured in the national media in September.)

Members of the USETI task force:
Edward Aboufadel (Mathematics, USETI Chair); Marty Abramson (Education); Philip Batty (Institutional Analysis); Katie Clark (Information Technology); Jon Jellema (Provost's Office); Laura Kapitula (Statistics); Brian Lakey (Psychology); Marie McKendall (Management); Christine Rener (FTLC); Neal Rogness (Statistics); Christine Smith (Psychology); Scott St. Louis (Student Senate).

Attachments: IAS form, validity report, other criteria report

---

[9] http://www.jcu.edu.au/learnandteach/public/groups/everyone/documents/how_to/jcu_121459.pdf
[10] http://www.stat.berkeley.edu/~stark/Preprints/evaluations14.pdf