TASK FORCE ON TEACHING EFFECTIVENESS FINAL REPORT

Georgia Institute of Technology

Task Force Members

Ayodeji Aladesanmi (undergraduate student, Econ)
Haigh Angell (undergraduate student, Econ and INTA)
Daniel Castro-Lacouture (Chair and Professor, School of Building Construction)
Lizanne DeStefano (Professor, School of Psychology, Director CEISMC)
Brian Fuchs (graduate student, ME)
Steven Girardot (Associate Vice Provost for Undergraduate Education)
Bob Kirkman (Associate Professor, School of Public Policy)
Joe Ludlum (Assistant Director for Academic Effectiveness)
Mirjana Milosevic-Brockett (Senior Academic Professional, School of Biological Sciences)
Loraine Phillips (Director of Office of Academic Effectiveness)
Ruth Poproski (Academic Professional, CTL – currently Associate Director of CTL at UGA)
Skanda Prasad Ponnathpur Nagendra (graduate student, ECE)
Narayan Shirolkar (graduate student, MSE)
Don Webster (Chair and Professor, School of Civil and Environmental Engineering)

Chair: Chair Bonnie Ferri (Vice Provost for Graduate Education and Faculty Affairs; Professor, School of Electrical and Computer Engineering). Bonnie.ferri@gatech.edu

Contents

Overview

The Task Force on Teaching Effectiveness was formed by the Faculty Executive Board (FEB) to examine the Course Instructor Opinion Survey (CIOS) instrument, alternative ways to measure of teaching effectiveness, and the use of student comments from CIOS. The members of the task force included students, tenure-track faculty, non-tenure track faculty, and administrators in order to ensure that a wide spectrum of perspectives were represented. The task forced started its work in December 2017 and concluded in April 2019 resulting in recommendations to the Faculty Executive Board for changes to processes and procedures and recommendations on best practices to the colleges, units, Offices of Academic Effectiveness and Faculty Affairs, the Center for Teaching and Learning, and SGA.

Charge of the Task Force on Teaching Effectiveness

The Faculty Executive Board specifically charged the Task Force to evaluate the CIOS in terms of

1. The instrument itself: what is being asked and how it is being implemented.
2. How it is being used formatively and summatively. Identify other possible assessment measures.
3. Releasing the comments to the School Chairs.

Each of these items was explored and is discussed in detail in the report below in the sections labeled Item 1 – Item 3 corresponding to the numbers in the list above.

History

This current task force was formed as a follow-up to a previous CIOS Task Force that was formed by the FEB in 2017 in response to a request by the Student Government Association that the student comments in the CIOS be released to School Chairs in an effort to improve teaching and to have the student voices heard. The first CIOS task force concentrated on the release of the CIOS comments while the second task force, the Task Force on Teaching Effectiveness which was active in 2018-2019, had the broader charge listed above. The major findings of the first task force and its recommendations include:

- There were strong opinions from the faculty regarding inappropriate student comments, especially those that are racist or sexist. As a result, SGA agreed to start a campaign for more civility among students in completing the CIOS.
- The faculty focus groups were almost evenly split in their support/opposition of releasing comments to School Chairs. When asked what would change their minds from opposition to support, the two most common suggestions were to address the sexist or racist comments and to limit the use of comments to situations when problems exist.
- A follow-on task force was recommended to explore the release of comments more thoroughly and to examine the CIOS more broadly and deeply.

Prior Georgia Tech studies and task forces examined related issues:

- 2003 College of Engineering Proposal for Peer Evaluation of Teaching – recommended peer evaluation
- 2008 and 2010 CIOS Task Forces – examined wording of questions and using an external vendor for implementing the CIOS

- 2017 Task Force on the Learning Environment – recommended mid-term evaluation, ongoing instructor development, recognition of the importance of the learning environment and mechanisms for addressing problems
- 2018 Path Forward – addressed issues of culture, student health and well being, inclusion and diversity

Methodology

In order to represent a broad spectrum of perspectives, this task force was composed of a mixture of tenure-track faculty, non-tenure-track faculty, administrators, undergraduate students, and graduate students representing five colleges. Their activities included doing a literature review on teaching effectiveness; statistical analysis of the CIOS; discussions with deans, associate deans, and school chairs; three campus-wide town halls; student input through the student members of the committee and through a presentation to SGA; and a pilot study of proposed changes to the CIOS. The data gathering of this task force was supplemented by the information gained from the faculty focus groups that were conducted by the first task force.

Major Findings

A brief overview of the major findings of this task force include the following points, which are discussed in more detail in the other sections of this report:

- The CIOS results show overwhelmingly positive student opinions of faculty teaching effectiveness
- Teaching effectiveness has dimensions in terms of *how we teach*, *what we teach*, and *what its impact is on long term student success*. The CIOS is an important but incomplete measure since it only addresses the dimensions of teaching effectiveness involving student perceptions of how we teach (ie, faculty-student engagement, inspiring interest, availability, etc.). What we teach (ie, course content, learning outcomes) is not represented in CIOS.
- There is an overreliance on CIOS as the sole measure of teaching effectiveness in faculty annual reviews, promotion, and tenure.
- There is a need for accountability of poor teaching

The exploration of the three items in the list of charges given by the FEB to the Task Force are discussed in detail in the sections below. While the release of all CIOS comments to School Chairs was the original SGA request, this recommendation is not being pursued. The main underlying reason for the request was that students wanted something to be done about poor teaching. In the discussion in Item 3 below, the recommendation is redirected to address the issue on accountability for poor teaching, where several sources of information including student comments should be used as part of the solution. It should be noted that many students assume that their comments are being viewed by chairs or individuals other than the instructor.

The Conclusion Section summarizes all of the recommendations of the Task Force, with the details of the recommendations given in the sections below. It should be noted that the American Sociological Association came out in September 2019 with a position statement that mirrors many of the recommendations of the Task Force. Their statement, included as Appendix D of this report, was endorsed by 17 other professional/scholarly organizations.

Item 1: Analysis of the CIOS Instrument

The CIOS was analyzed to examine the following questions:

- What does CIOS tell us about student opinions on teaching at Georgia Tech?
- Is there bias in the results?
- Are the CIOS results correlated with high class GPAs?
- Can the validation of the instrument be determined?
- Are there better ways of wording questions, and can some redundancy be removed in order to shorten the survey?

The single most commonly reported response on the CIOS is for the question "Considering everything, the instructor was an effective teacher." The response to this question is used heavily in all promotion and tenure deliberations, annual faculty reviews, as well as an indicator of possible problems with an instructor or course. Figures 1-3 show the distribution of the responses for this question for all course sections taught in 2018 on the Atlanta campus that had at least five respondents. Figure 1 is the distribution for all course sections, both graduate and undergraduate. This data represents over 6000 course sections, and the average response rate was 57%. This data paints a very positive picture on how students rate instructors - over 50% of section instructors have a CIOS rating of 4.7 or higher on this question.

Figure 2 shows the results of the "effective teacher" question for only undergraduate courses, separating them into sections with enrollments of 50 or larger and sections with under 50 students in order to examine the effect of class size. There is a slight shift in the results, with larger classes being rated slightly lower, but both types of classes still have distributions that peak in the range of 4.7-5.0.

Another way of examining the 2018 data is to look at the raw responses in those course sections, without aggregating them into sections. Figure 3 shows a frequency plot of over 100,000 student responses of the integral choices 1-5. Again, this plot shows the same trends in responses with the highest selection being the choice of 5. Note that there are 18% of responses that are ≤3.

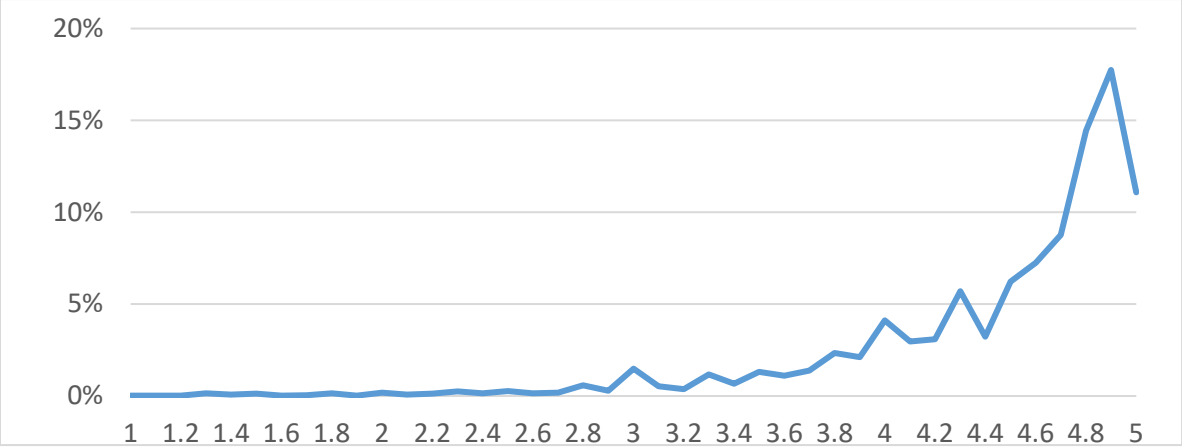

Figure 1: Distribution of responses for the "effective teacher" question for all course sections (with at least 5 responses) taught on the Georgia Tech Atlanta campus during calendar year 2018.

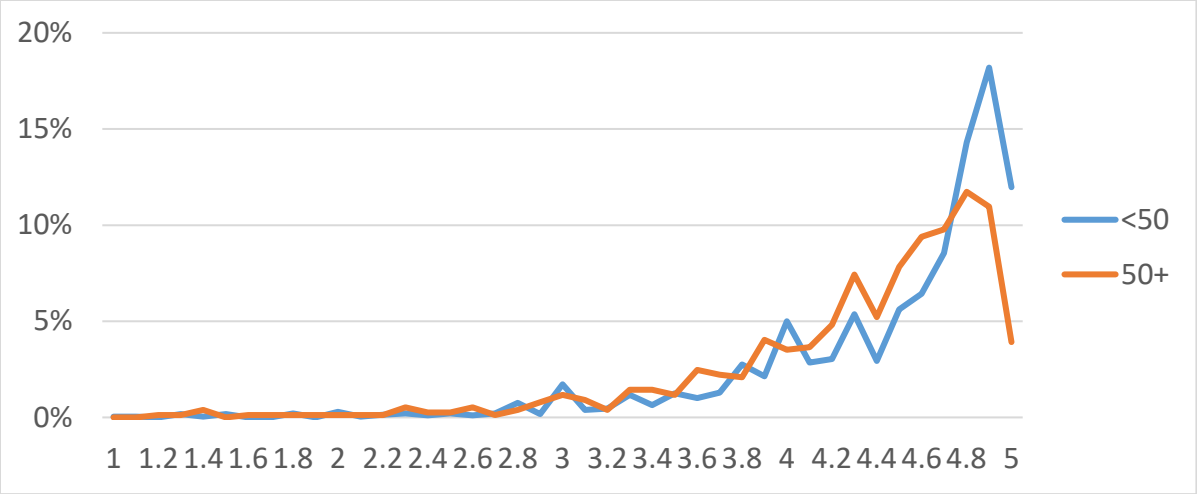

Figure 2: Distribution of responses for the "effective teacher" question for all undergraduate course sections, separated by large class enrollments (50+) and small class enrollments (<50).

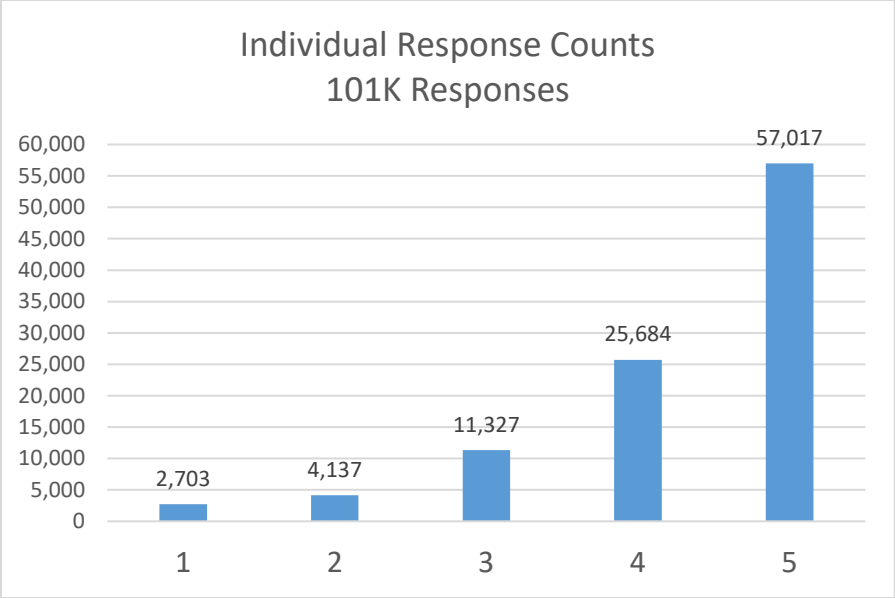

Figure 3: Total numbers of student responses in 2018 for each of the possible ratings, 1-5, on the "effective teacher" question

While the vast majority of section instructors receive good student responses on this CIOS question, there is a tail end of the curve at the low end. Table 1 separates the data from Figure 1 into colleges. The last column shows the results for the Atlanta campus as a whole. For example, 4.7% of instructors received a rating of 3.0 or lower for the "effective teacher" question.

Table 1: Summary of the results of the CIOS question "Considering Everything, this is an effective teacher." The percent of all sections of courses in 2018 that had responses of 3.0 or lower and 3.5 or lower (on a range of 1-5) are listed by college and for the Institute as a whole.

	COC	COD	COE	COS	IAC	SCOB	GT
CIOS ≤ 3.0	3.9%	3.4%	8.3%	3.4%	2.6%	3.7%	4.7%
CIOS ≤ 3.5	8.6%	10.2%	13.2%	7.9%	4.5%	6.9%	8.6%

There may be some legitimate reasons for low CIOS ratings – perhaps someone is trying something new in their courses and had problems. We don't want to discourage pedagogical experimentation, but the low ratings warrant further review. One poor instructor casts a very large shadow on the experiences of students.

Bias: There have been a number of studies that report bias existing in student opinion surveys, see for example [MacNeil, 205; Mitchell, 2018; Peterson, 2019; Weinkle, 2019]. We did not examine bias directly in the Georgia Tech CIOS results. Instead, we examined the differences in the individual responses of the "effective teacher" question between women and men instructors in all sections taught in 2017. This was tested through analysis of variance (ANOVA), using Instructor Gender as our primary factor. To control for known differences, College and Class size categories were also included as separate factors. By drawing off the effects of these items and interactions, the specific impact of gender can be isolated.

The large number of responses available provided a substantial amount of validity to the analysis, resulting in significant effects for all factors and several interactions. We see a slight but statistically significant difference for gender overall, so there may be some level of gender bias in the CIOS results.

Table 2 shows those gender differences separated by colleges. The colleges vary in the degree of difference, with Engineering, Business, and Sciences having small, not statistically significant differences. These small differences between the genders on this particular question would have very little, if any, impact on promotion and tenure cases, which use a larger granularity when examining cases. However, this difference points out the need to educate committees and reviewers on how to interpret CIOS results. A good reference is [Linse, 2017], which provides a guide for administrators on how to interpret the results of student opinion surveys. A more recent paper, [Peterson, 2019], reports on a study where making students aware of potential implicit bias right before they complete a survey helps to remove the bias in the results.

Table 2: Differences between student responses on the "effective teacher" question for all course sections in 2017, separated by gender and by college.

	COC	COD	COE	COS	IAC	SCOB	GT
Women	4.180	4.404	4.200	4.240	4.332	4.379	4.285
Men	4.313	4.547	4.237	4.231	4.472	4.407	4.366
Gap (Women – Men)	-0.133*	-0.142*	-0.037	0.010	-0.140*	-0.028	-0.081*

* - significant at $p < .05$

There are common strategies to mitigate potential bias in qualitative evaluations used in personnel decisions such as hiring and promotion, and we recommend using those methods with student evaluations:

> *All administrators and review committees who will use the CIOS results for evaluation of faculty should be trained on appropriate usage of the information, including implicit bias training, and the CIOS should be examined to see if modifications such as those discussed in [Peterson, 2019] have an impact on bias.*

While the potential gender bias in CIOS is not expected to have a critical effect on promotion and tenure, the bias might have a much larger impact on teaching awards. Recall that 50% of the CIOS ratings for the "effective teacher" question are in the range of 4.7-5.0, so that a difference of 0.08 between genders is quite relevant in this range. This task force asked CTL to revise their CIOS awards, which had been based solely on receiving the highest CIOS scores on this question.

> *All teaching awards should consider this potential bias when evaluating candidates.*

Effect of GPA: There is a common belief that one way to get better CIOS ratings is to make the course easy or to give high grades. In the CIOS analysis, neither of these findings were proven to be true. Figure 4 shows a scatter plot of the average of the final grades for each class ("ClassGPA") versus the CIOS Interpolated mean for the "effective teacher" question (ie, the value that is reported to instructors). A strong correlation between these values would be a more narrow grouping of points around a straight line. A wide scatter indicates very little correlation. In particular, the correlation coefficient is 0.275 using the raw data or 0.22 when taking into account prior GPA of individual students. The shared variance is 7.6%, that is, about 7.6% of the variation in overall grades for a course fall in line with the between-course variation of "effective teacher" ratings.

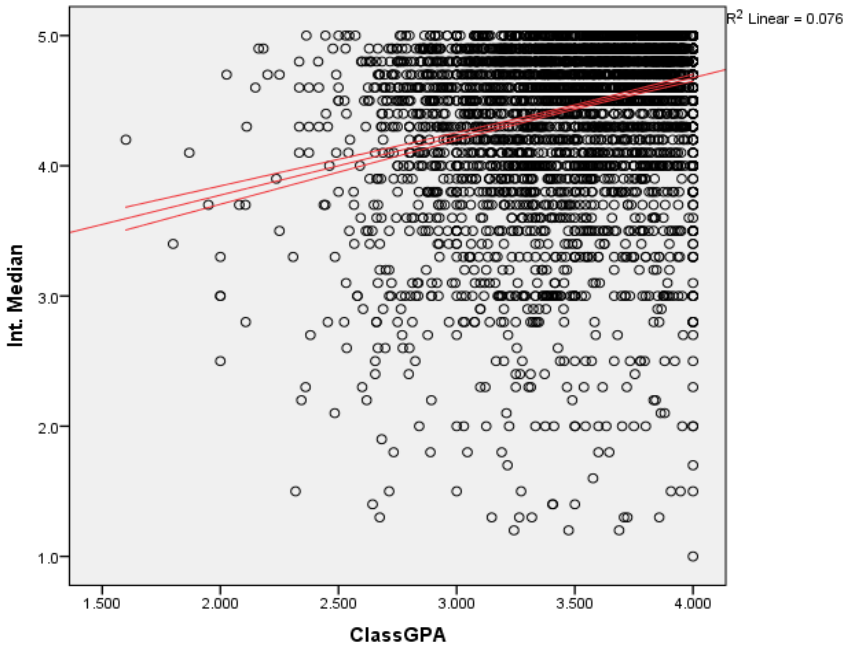

Figure 4: Scatter plot of section GPAs versus CIOS response on the "effective teacher" question for all sections (with at least 5 responses) in 2018. The red lines are linear regression results.

It is important to note that while we do find a small association between grades and the overall effectiveness question, it does not address the question of causation. While it is possible the students' expected grades in the class drive the teaching effectiveness ratings, another explanation could be that the higher ratings of teaching reflect a more effective instructor, which results in better academic performance on the part of the student. To test the first possibility, that is, the one most favoring the "easy course" theory, ratings need to be compared to a measure of course difficulty.

To explore this, the data for overall teaching effectiveness used above was correlated with the average number of hours per week spent on a course, using this measure of effort as a proxy for the difficulty of a course. If an easy instructor is more favorably rated, then as effort goes down, instructor ratings go up – a negative correlation. The correlation coefficient for these two items was at 0.004, showing no correlation between the CIOS rating and the number of hours spent on the course. As such, it appears that the difficulty of a course is *not* associated with the CIOS rating for overall teaching effectiveness.

Validity: The CIOS contains two categories of questions, one on the teaching and one on the course. All of the questions that relate to teaching correlate highly with the "effective teacher" question. Similarly, all of the questions that relate to the course correlate highly to the "effective course" question, indicating that the survey is internally consistent. For example, there are strong correlations between the "effective teacher" question and the questions about the instructor characteristics identified in Table 3.

Table 3: Instructor characteristics found to correlate strongly with the "effective teacher" question.

Instructor's availability for consultation
Instructor's clarity on the course material
Instructor's enthusiasm for the course
Instructor's respect and concern for students
The instructor communicates what it would take to succeed in the course
The instructor stimulates interest
The helpfulness of feedback on assignments

Wording and Length: The Task Force examined each of the questions on the CIOS for ambiguity, for redundancy, and to determine if there is missing data not being asked. They also wanted to address recommendations given by the Task Force on the Learning Environment and the Path Forward Initiatives on inclusion.

The Task Force is making specific recommendations to the CIOS that are shown in Appendix A that modify wording to improve clarity, reduce redundancy of the open comment questions, and adds a question on the learning environment:

"The instructor provided a positive environment in which to learn, regardless of my race, color, religion, sex, national origin, age, disability, sexual orientation, gender identity, or veteran status."

A pilot study was done in the summer of 2018 with these changes, as well as an additional question that addressed the comments. See the section later in this report that discusses the pilot study.

Recommendations for Item 1

Including the recommendations given above to reduce bias, below is a summary of the recommendations of the Task Force for modifying the CIOS in response to the exploration of Item 1 of the task force's charge:

- Modify the CIOS as shown in Appendix A to improve clarity of questions, remove redundancy, and add the question on the learning environment.
- Require implicit bias training on the use of student opinion surveys for anyone who uses the CIOS in faculty evaluation, including evaluation for teaching awards.
- Investigate modifications to the CIOS, such as those studied in [Peterson, 2019], that have the potential to reduce the bias exhibited by students who complete the CIOS.

Item 2: Assessment of Teaching Effectiveness

Quite a bit of the Task Force's effort concentrated on determining what is meant by "teaching effectiveness" and what are metrics of teaching effectiveness in addition to student opinion surveys. The other part of this charge was to determine how CIOS is being used at Georgia Tech.

Teaching effectiveness is the single-most studied topic in education, with thousands of published articles and several books and journals devoted to it. A common interpretation of teaching effectiveness is that the teaching was effective if the resulting student performance in the class was good. A common way to measure teaching effectiveness is to administer end-of-term student opinion surveys, such as CIOS, that are student evaluations of teaching (SET). There is a great deal of debate whether SETs, such as CIOS, correlate to student performance. Two meta-analyses of the relationship between SETs and learning outcomes found a low-to-moderate correlation between the two [Cohen, 1980; Feldman, 1989]. Studies included in those meta-analyses focused on multisection studies, that is, instances in which a number of sections of a given course are taught by different instructors using the same syllabus and the same exams.

A recent article that performs a meta study of many of the earlier published results casts doubt on this correlation [Uttl, 2017]. This recent study subjects the meta-analyses of the previous papers to rigorous scrutiny and re-ran the analyses using a methodology that could account for the effects of small sample sizes and publication bias; they also conducted a more comprehensive meta-analysis of their own. This particular study found that previous studies indicate a weak-to-moderate correlation between student performance and student ratings of instructor when the number of sections, N, of the course was small (<10). The correlation dropped when the number of sections for a course was increased, with N=194 as the largest number of course sections studied. (There was no explanation in this reference on how a common test or common performance evaluation was done across such large number of sections.) The

paper contends that once the effects of sample size, that is, the number of course sections N in a study, and publication bias are taken into account, any perceived correlation between SET and student learning washes out.

A simple example demonstrates the limitations of student opinion surveys:

> Consider two instructors who are teaching the same course, and both receive a 4.5 score on the CIOS "effective teacher" score. One instructor covers the entire syllabus and meets all the learning objectives. The other instructor only covers ¾ of the syllabus and does not meet the learning objectives. Which is the more effective teacher? If student opinions are the only metrics, then they would appear to be equal.

If student opinion surveys are questionable as a good measure of student learning, what do they measure? The findings of the CIOS statistical analysis discussed in the previous section show a strong correlation between the student response to the question "Considering everything, the instructor was an effective teacher" to the other questions on CIOS that revolve more around how the instructor engages with the students, specifically, those instructor characteristics listed in Table 3.

The Task Force did a literature survey to examine alternatives to student surveys as means of measuring teaching effectiveness, see the report on the literature survey in Appendix E. The report identifies fifteen different ways of measuring teaching effectiveness and gives an analysis on how and why each method might be used and gives a summary of research studies on those methods. The report in Appendix E is partitioned into four categories of measures: *student perspectives, peer review of teaching, instructor reflection and review,* and *evidence of student learning*. The literature supports that student evaluations of teaching are important but not sufficient. The consensus is that the assessment of teaching effectiveness is most accurate with a combination of measurements that come from the four categories. As a side note, we would like to recommend the use of *mid-semester student evaluations*, as discussed in the literature survey, these can be a very effective way to improve teaching and to improve end-of-term evaluations because the instructor has time to react to the feedback and make changes. The Task Force on the Learning Environment also recommended mid-semester evaluations.

In addition to archival studies on the topic, there is a national debate on the use of student evaluations of teaching, especially with regard to bias. For example, the Chronicle of Higher Education reported in January 2019 that many universities are grappling with the opposing perspectives that student evaluations of teaching are flawed in that students are not experts in education and they may have biased opinions, yet students do give valuable information about the instruction and the course [Doerer, 2019]. This article discusses approaches that different universities are taking, including de-emphasizing student evaluations of teaching in faculty evaluations and augmenting them with peer evaluation and self-evaluations; educating students and faculty about the potential for bias; and removing the "effective teacher" question while leaving the more specific questions on how the course was conducted.

More notably, the American Sociological Association came out with a position statement discussing the limitations of student evaluations of teaching and stating that "their use in personnel decisions is problematic". They recommend that institutions do not use student evaluations of teaching as the sole metric of teaching effectiveness, that evaluators for personnel decisions have training on interpreting

student evaluations, and that the trends in scores over time of an individual is more important than an absolute comparison among others instructors. The position statement is contained in Appendix D.

CIOS Use at Georgia Tech: Student opinion surveys, CIOS, are used commonly as a means of measuring teaching effectiveness at Georgia Tech. The response to the question "Considering everything, the instructor is an effective teacher" is the single-most quoted metric of teaching effectiveness and is required in annual reviews, periodic peer review, teaching award evaluations, and in all promotion and tenure documentation. In many cases, this information is the only required evidence of effective teaching. The major exceptions are the following instances where additional evidence of teaching effectiveness is required.

- Ivan Allen College requires committee-based peer review of teaching that are included in faculty evaluations
- The College of Science has a Director of Teaching Effectiveness (DOTE) to do peer evaluation of teaching, which is included in all faculty evaluations
- Lecturers across the Institute are required to supply a teaching portfolio as part of their promotion package. This teaching portfolio includes a reflection and self-evaluation of teaching, peer evaluation, the CIOS results, and sample teaching materials. A complete description of the teaching portfolio requirements is given in Appendix B.

Considering the overwhelming evidence that student opinion surveys are an important but insufficient way to measure teaching effectiveness, the Task Force feels that:

> *There is an overreliance on CIOS as the sole measure of teaching effectiveness in annual reviews, promotion, and tenure at Georgia Tech.*

The term "opinion" should be emphasized in the title of CIOS, Course-Instructor Opinion Survey.

Dimensions of Teaching Effectiveness: The Task Force felt that the reason that there is no single good measure for teaching effectiveness is because teaching effectiveness has multiple dimensions that can be described by *how we teach, what we teach, and what are the long term impacts of our teaching*.

How we Teach:

- What is the level of engagement between the instructor and student: clarity of instruction, availability of the instructor for additional help, responsiveness of the instructor, etc.
- Is the learning environment conducive for all students in the course to learn the material: inclusivity, civility, respect, course format and organization, etc.
- Does the instructor use evidence-based pedagogical methods: active learning, etc.

What we Teach:

- Are the course learning objectives and outcomes being met?
- Is the required course material being covered?
- Is the grading appropriate for the material and level of accomplishments of the students?
- Is the material made relevant to other courses or to real applications?

- Is there an acceptable level of consistently between course sections on material and grading?

What are the Long Term Impacts:

- Does this course facilitate student success in future courses and career?
- Does this course help to meet the GT mission – "we will be leaders in influencing major technological, social, and policy decisions that address critical global challenges"?

Another aspect of good teachers is that they have a process in place to consistently improve their teaching by using feedback from students, peers, literature, and their own reflection. They should also keep up to date on new developments in their discipline and in pedagogy in their courses.

Measures of Teaching Effectiveness Across Dimensions: No single measure of teaching effectiveness addresses all three dimensions of effective teaching. The literature survey in Appendix E lists many of the typical measurements, which are grouped into student-oriented measurements, peer evaluations, and self-evaluations. Table 4 gives a list of several typical measures of teaching effectiveness found from the literature review in Appendix E and adds program-level measurements. This table shows the Task Force's consensus of how the different measurements map to the different dimensions of teaching. The measurements are grouped as student, peer, self, and program oriented. Note that the effectiveness of some of these measures relies on how well they are done. Ideally, there should be a combination of measurements that combined adequately spans the three dimensions of teaching effectiveness. For example, peer observation of teaching is much more effective if the reviewers are trained and the assessment method is validated.

While students may be able to evaluate the learning environment and the student-faculty engagement, they are not good judges to determine if the learning outcomes are met since they are not experts in that discipline. Evaluation by peers does provide that expertise. Peer review of course materials adds a different perspective than classroom observation and is a strong indicator of "What we Teach". More generally speaking, "peer evaluation" may also be interpreted as peer oversight, for example a multi-section course might have a course coordinator to ensure that the material being covered is consistent across sections and up-to-date and that there is some consistency and appropriateness on how grades are assigned. Or, the course instructors can meet once a term and compare course coverage and materials (exams, projects) to encourage coordination. Peer observation of teaching can be very effective if the observer is well-trained in pedagogy.

Teaching portfolios, required by many universities as part of promotion packages, is a single document that includes several of these measurements, most commonly: self reflection and evaluation, student evaluations, evidence of student learning, peer evaluations, and sample instructional materials. Georgia Tech does require a teaching portfolio for the promotion of Lecturers, as shown in Appendix B, but not for other academic faculty.

A major finding of this Task Force is

> *If the goal is to have a complete assessment of teaching, then multiple measures should be used in combination that span multiple dimensions of teaching effectiveness.*

Table 4: Mapping of measurement methods to dimensions of teaching effectiveness. The table is coded so that good indicators of that dimension are shown by a large check mark (✔) , and mediocre indicators are shown with a small check mark (✔). The absence of any mark indicates very little confidence in that measure for that dimension.

	Dimensions of Teaching		
Ways of Measuring Teaching Effectiveness	**HOW WE TEACH** (engagement, learning environment, teaching methods)	**WHAT WE TEACH** (learning objectives, coverage, grading)	**LONG TERM IMPACT** (future success, GT mission)
Student Opinion Surveys (CIOS)	✔		
Focus Groups with Students in the course (done by a trained facilitator)	✔		
Evidence of Student Learning (performance on a standardized test or on common exams across sections of the same course)		✔	
Student Longitudinal Evaluation (success in follow-on courses, longitudinal surveys)		✔	✔
Peer Observation* (including faculty from CTL)	✔	✔	
Peer Review of Materials (faculty from the same discipline)	✔	✔	
Self Reflection & Evaluation	✔	✔	✔
Program or Institute metrics (comparative grading, presence/absence of student complaints, DFW rates)	✔	✔	✔

*The effectiveness of observation is influenced by the training of the reviewer, the validity of the review process, and the number of times that the review is being done within the term.

Program-Level Metrics Related to Teaching Effectiveness: Table 4 includes some Program or Institute metrics in order to represent a *program-level perspective on teaching effectiveness*: pockets of poor or inconsistent teaching impacts the overall quality of a program and the students' experience in that program and at Georgia Tech. Some program-level measurements that are related to the quality of teaching include the success in follow-on courses, exit and alumni surveys, grade distributions in the

course section compared to other sections of the same course, and DFW rates for the course (DFW rates correspond to the number of students who receive a D, F, or W in the course).

To demonstrate the relationship of these metrics on teaching effectiveness, we looked at the DFW rates for some selected required undergraduate courses that are taught across multiple sections. Figure 5 gives four specific courses taught on the Atlanta campus during the time period from Spring 2018 through Spring 2019 and shows the two extremes of DFW rates for different sections of those courses, the highest and the lowest rates. In Course "2", for example, one section had a DFW rate of 50% while another section had a DFW rate of 1.8% (during the same term). There are two issues that arise from this data:

- High DFW rates do have implications on the teaching effectiveness in that course since a high DFW rate means that a large number of students are not succeeding in that section of that course. A large DFW rate may indicate that there is a problem with the course, with the preparation of the students for the course, or with the instruction in the course.
- A large difference of DFW rates or grades among course sections means that students in different sections are not getting the same experience with that course. Consistency in instruction impacts the overall quality of the program. An outlier of a very high DFW rate for one section compared to others indicates that the cause is less likely to be the course itself or the preparation of students for the course, and it is more likely to be a problem with the instruction.

There are similar issues with consistency in teaching across course sections in terms of grading strategies and the course coverage. For example, in a recent review of a course at Georgia Tech it was found that there was a 50% mismatch in coverage between two sections of a required course. Regardless of the CIOS ratings for that course, the answer to the question of whether that course was being taught effectively would have to be "No".

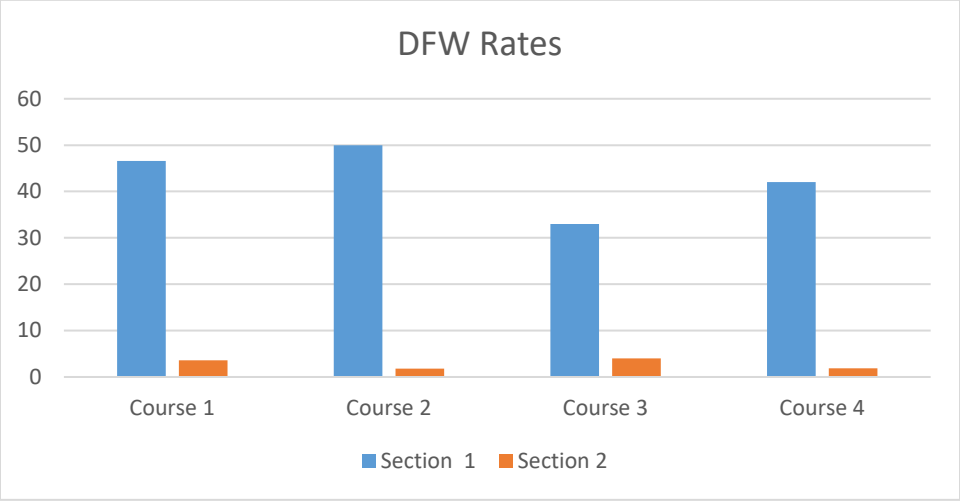

Figure 5: Four selected multi-section required undergraduate courses taught on the Atlanta campus, showing the two extreme DFW rates for each course showing the highest and the lowest rates among the different sections of the course.

Recommendations for Item 2

The recommendations of this Task Force incorporate research on the subject of teaching effectiveness and, as mentioned previously, mirror many of the actions being taken at other universities.

- There is an overreliance of CIOS numerical scores to measure teaching effectiveness at Georgia Tech. Schools/Colleges/Institute should realize what the CIOS is and isn't – it is a student opinion of one dimension of teaching: *how we teach*. Teaching effectiveness is multi-dimensional (how, what, long term impact): it needs multiple ways to measure it if you want to span the different dimensions and have an accurate assessment.
    - *Promotion, annual reviews, periodic peer reviews:* Two colleges, IAC and COS, already use peer evaluation and CIOS. Promotion for Lecturers requires a Teaching Portfolio. A minimal step would be to augment annual reviews and promotion packages with a reflection and self-evaluation statement similar to the one that is currently required as part of the teaching portfolio for Lecturers (see the description of this item contained in Appendix B).
    - *Accountability:* If there are indications that a problem with teaching is present, then an investigation of the problem should include multiple types of measurements to better understand the scope and validity of the problem, not rely on a single metric.
- Units should encourage the use of mid-semester evaluation for formative assessment.
- Programs should consider teaching effectiveness on a broader scale, not just tied to a particular instructor in a particular course. Program-level metrics and peer evaluation/oversight can be used to examine and enforce some quality control over teaching in a unit in terms of quality of instruction and consistency of coverage and grading among different sections of the same course.

Item 3: The Use of Comments

Student comments on the CIOS give context to the quantitative data and hold valuable information about the course and about the instruction. The SGA first requested that the comments be released to the school chairs as a means to identify problems in teaching and to address those problems. The SGA cited a number of peer institutions that do release the comments to the school chairs, see Table 5.

Table 5: Peer institutions that allow department heads to view the comments on the student evaluations of teaching opinion surveys.

Cal Tech	UT Austin
Carnegie Mellon	MIT
Cornell	Penn State
UC Berkeley	Purdue
U Michigan Ann Arbor	Virginia Tech

The first CIOS Task Force in 2017 conducted focus groups in five colleges (CoC was the exception) including 119 faculty members on the issue of releasing comments to school chairs. A mock ballot on releasing versus not releasing the comments resulted in 48 in favor, 41 opposed, and 23 unsure. When asked what might change your vote to favoring, 17 people said addressing racist or sexist comments, 10 people said limiting the situation to when a problem exists, and 5 people said to release a summary prepared by someone trained in processing qualitative data. That CIOS Task Force recommended that a campaign be started to improve civility among students to provide constructive feedback, and the CIOS instructions were modified to this same effect.

When pushed further on why students felt that releasing the student comments was so important, an SGA officer said "Because there are some horrible teachers out there, and no one is doing anything about it." This message indicates that the issue is not really about releasing comments, rather about providing *accountability for poor teaching*.

> *While the first CIOS Task Force addressed the issue of releasing comments, the Task Force on Teaching Effectiveness focused more on accountability, with student comments being part of that solution.*

The need for accountability resonated with administrators. Discussions with school chairs, associate deans for academics, and deans provided unanimous support for the release of the comments. Their perspective was to use the comments for improving teaching in their schools and for overall program improvement. One school chair said that accessing the CIOS comments would help to identify if a low CIOS rating was due to the course, the instruction, or the students' preparation. Another said that they would like to compare comments associated with low CIOS ratings with those associated with high CIOS ratings to identify what it takes to be successful.

One of the most challenging aspects of academic administration is handling student complaints about teaching. One or two student complaints about an instructor might be outliers, or they might represent the opinion of the entire class. Without being able to access the comments for that class to determine if the complaints represent a larger pattern of behavior, those one or two complaints take on a role of increased importance.

> A sample existing scenario demonstrates the role of comments in addressing problems: Some students complain about Professor A's teaching. Professor A claims those complaints are unfounded and points to a high rating of the "effective teacher" question in defense.
>
> It may turn out that Professor A has one unacceptable teaching practice (such as very long delays in returning graded work, lack of responsiveness, etc) but is good otherwise.
>
> It also may turn out that Professor A has a learning environment that the majority of students find acceptable, but that a minority of students find very objectionable. The resulting CIOS numerical scores are skewed by the majority opinion, but this does not mean that the minority group of students do not have a valid concern.

If school administrators can access the comments, then they can assess the extent of the problem within the context of the rest of the course and coach instructors on possible targeted actions.

The Task Force considered the views of the community when coming up with a recommendation on CIOS comments. Students and administrators are overwhelmingly in favor of allowing administrators access to comments for the purpose of using the information to help address instances of poor teaching and for improving the overall quality of teaching in a school or program. Faculty in the focus groups were split in their views (48 in favor, 41 opposed, 23 unsure) with major concerns being how the comments might be used or misused in faculty evaluations and the presence of inappropriate comments, especially those that are sexist or racist. The Task Force accepted the recommended modifications suggested by 32 of the "opposed" and "unsure" faculty in the focus groups by limiting the use of the comments to when there are problems, ensuring the reviewers of the comments are properly trained in handling qualitative and potentially biased data, and directly addressing inappropriate comments.

The Task Force decided to test the use of one question in the Summer 2018 pilots:

> "Is there anything that you would like to share with both the instructor and academic administrators about this course?"

The student comments in that pilot study contained information about the course content (48.7%) and the instruction (39.5%), with the majority of comments providing constructive feedback. The comments give context to the numerical data. This information could help school administrators improve the overall quality of their programs. There is a detailed discussion of this pilot later in this report.

Recommendations for Item 3

- Develop processes to enforce accountability for poor teaching, starting with coaching for improvement but including stronger measures if needed. The Task Force on the Learning Environment made a similar recommendation in terms of addressing problems with the learning environment. As mentioned in the previous section, "poor teaching" needs to be determined using multiple measures of teaching effectiveness such as those defined in Table 4.
- Allow administrators to have access to CIOS comments, as needed, as part of an investigation on a potential problem with a particular course or an instructor.
    - Indications of a problem might come from any of a number of sources such as low CIOS scores, student complaints, outlier DFW rates or class GPAs, survey results, low success in follow-on courses, etc. The goal of the investigation is to pull together different sources of information, including student CIOS comments, to determine the extent of the problem and the context in which it arises, and to help guide solutions.
    - Require administrators who need to access comments be trained on how to interpret and use qualitative data, especially data with potentially biased comments. Rather than quote individual comments or place undue emphasis on them, a common practice in interpreting qualitative data that we recommend is to find common themes among the comments and then count the number of comments in each those themes to assess the importance of each theme. Additional guidelines for training of administrators on the use of student opinion surveys are in [Linse,2017].
- Address faculty concerns about racist, sexist, and abusive student comments, especially those that could potentially violate Georgia Tech's anti-discrimination policy. Such actions might include adding a statement to the CIOS that makes students aware of bias, a practice that was reported in a recent article to have resulted in a reduced bias [Peterson, 2019]. Another

possibility is to allow instructors to report comments that potentially violate this policy. For example, if a committee determines that the comments violate this policy, then those comments and the associated quantitative data would be removed from the CIOS record of that course. In this process, students would maintain free speech in making their comments, but there is no requirement that the Institute need use inappropriate comments in their evaluation of teaching.

- Provide students with opportunities to give feedback outside of CIOS, such as using a webform on a unit's home site, student focus groups, and targeted student surveys. These mechanisms can broaden the topics from just a particular instructor or course, to cover areas such as academic advising, the curriculum, extra-curricular opportunities, research advising, etc. For example, a webform can collect information similar to that of student exit surveys (given to graduating students) but is much more timely and open to all students. It should be noted that this sort of feedback form was presented during an SGA Senate meeting, and students felt that this mechanism for obtaining student feedback was not as good as using CIOS comments since the webform would have far fewer responses and would be require students to duplicate their efforts from the CIOS. A small pilot study of using webforms is described in the next section.

Results of Pilot Studies

Two pilot studies were held in 2018, one that examined modified CIOS questions and the other that examined the benefits of using a webform to obtain student comments outside of CIOS. Both are described below.

Summer 2018 CIOS Pilot Study

A pilot study was completed in the Summer 2018 semester to investigate two versions of a modified CIOS given in Appendix C. The differences between the current version and the modified versions are shown in Table 6. The two modified versions used in the pilot included questions that were modified for to improve clarity, reduction of redundant questions, and a new question on inclusivity of the learning environment. The difference between the two piloted versions is the addition of one question for one of the pilots:

> "Is there anything that you would like to share with both the instructor and academic administrators about this course?"

Instructors from four colleges elected to participate, providing 42 courses for the pilot group. A control group of 158 courses was selected from Summer 2018 CIOS results, based on being comparable to pilot group in terms of represented colleges and departments, class size, and course levels. For this analysis, individual responses were used instead of course composites, providing a total of 2,228 responses (609 in the pilot group, 1,619 in control group). The study group traits are presented in Table 2 of Appendix C.

The primary concern in changing the instrument is the possibility that the altered question prompts and additional content might affect the scores of the current items. "Overall effectiveness of the instructor" is of particular concern as it is used for teaching evaluations in faculty Promotion and Tenure review, and the Class of 1940 Award. To be thorough, all course and instructor ratings were tested.

Because several factors are known to associate with CIOS scores, an Analysis of Covariance (ANCOVA) was used to test for differences. In addition to the study group, traits with known associations with CIOS

scores were added as controlling variables (college, sex of instructor) and as covariates (course level, course size). The full pilot study report is presented in Appendix C. With other factors accounted for, only a single item showed a significant difference between the pilot and control groups: The degree to which activities and assignments facilitated learning, $F(1,2168) = 3.929$, $p = .048$, partial $\eta^2 = 0.002$.

Table 6. Differences between Current and Pilot Versions of CIOS.

Current Version	Pilot Version
Course Items	
Degree to which activities and assignments facilitated learning:	Degree to which exams, quizzes, homework (or other evaluated assignments) measured your knowledge and understanding:
What was the best aspect?	What were the best features of the course, such as lectures, activities, assignments, and projects?
How could it be improved?	How could this course be improved?
Instructor Items	
New Item	The instructor provided a positive environment in which to learn, regardless of my race, color, religion, sex, national origin, age, disability, sexual orientation, gender identity, or veteran status.
What was the greatest strength?	What was the greatest strength of the instructor?
What is the most needed improvement?	How can instruction be improved?

Administrator-Shared Comments: For the participants presented with the opportunity to provide comments available to department chairs or administrators, a total of 76 CIOS participants (20.1% of the Pilot v1 group) provided feedback. The percentage of students who chose to comment to an administrator is comparable to the percentage that normally make comments.

Most of the respondents did comment on course content (48.7%) or about the instructors (39.5%). The majority of responses (63.2%) also provided constructive feedback: highlighting specific aspects of the course or instructor efforts that were effective, or did not work. Many of these comments included suggestions for improvements.

Conclusions: Based on the available data, the survey updates and revisions do not appear to impact scores on the other CIOS items. Similarly, the new inclusivity survey item did not appear to impact scores on the other CIOS items. The Administrator-shared comments did garner information that could be used by department chairs. Should the Office of Academic Effectiveness be so directed, any updates proposed by the Task Force could be implemented.

Student Feedback Webform Pilot

As a means of giving students an avenue outside of CIOS to voice their opinions, the Task Force created a student webform and piloted in the Schools of Mechanical Engineering and Civil and Environmental Engineering. In ME, the survey link was sent out in emails to students while the CEE survey link was both sent to students in and email and put on a school website starting in October 2018, which yielded a higher response rate (24 for CEE versus 4 for ME). The introductory text for the two webforms is very similar; the text for CEE reads:

> Please use this form to provide comments and suggestions about your experience in CEE. Your comments will be reviewed by academic administrators within the School, with the goal of improving your experience, as well as the experience of current and future students.
>
> Your feedback will be sent directly to Dr. Susan Burns, Associate Chair for Administration and Finance. You can choose to remain anonymous, or to include contact information. Please understand that if your identity remains confidential, Dr. Burns will follow up on the issue, but she not be able to communicate with you regarding resolution of the issue.
>
> Your responses are completely anonymous unless you choose to identify yourself. If you do identify yourself, we will follow up with a response and/or resolution.

To date, they have received 24 responses, although significantly more people have viewed the survey without entering feedback. In general, the responses can be sorted into three main categories:

- **Curriculum**: The majority of curriculum feedback (8/14 comments) focused on changes to the Environmental Engineering degree, with requests centering on 1) the addition of a graphics course, 2) a change in required laboratory, and 3) a change in required engineering mechanics courses.
- **Faculty**: The majority of feedback (8/9) focused on two faculty members.
- **Advising**: One complaint was filed about academic advising in the School.

It is important to note that there have been no acute problems filed through this mechanism to date, such as “my instructor cut our exam off after an hour”, or “my instructor is not following the syllabus”. Most of the feedback is anonymous (17/24), but in the cases where an email address is given, either Susan Burns, or the relevant group leader has followed up with the student.

When feedback is submitted, the survey results are sent to Susan Burns (Associate Chair for Admin and Finance), who removes identifying information and forwards the issue to the relevant faculty group leader, administrator, or curriculum committee. Some problems are longer term, such as faculty issues, and are managed by Susan Burns and Don Webster (School Chair).

The curriculum issues summited in the form are being considered by the Environmental Engineering curriculum committee and the CEE Undergraduate Committee. Mentoring of faculty is underway for faculty-specific complaints.

Observations: the student feedback submitted in this webform is on broader topics than that provided in CIOS, and is useful for that purpose alone. Note that the response rate is much lower than found with

CIOS, so it represents a much smaller sample of student opinion on individual courses. This webform was discussed in an SGA meeting, and students mentioned that they feel that the CIOS is an easier mechanism for them to convey opinions to the school administrators since they are already filling them out for the instructors.

Conclusion

Appropriate use of student evaluations of teaching effectiveness has been debated hotly in the national press and in scholarly articles during the last few years. This Task Force was charged by the FEB with evaluating CIOS with regard to 1. the instrument itself, 2. how it is being used and if there are other assessment methods for teaching effectiveness, and 3. releasing the comments to school chairs. The Task Force reviewed the literature on teaching effectiveness, performed data analysis on the CIOS instrument and on its results, examined the use of CIOS comments, did pilot studies, and examined other metrics of teaching effectiveness. They also gathered input from students, faculty, and administrators at Georgia Tech through faculty focus groups, town halls, and individual and groups discussions. Each of the three main charges to the Task Force, identified numerically above, resulted in recommendations that are described in detail in each of the corresponding sections of this report.

Some of the most critical action items are ones that impact Georgia Tech's mission and personnel decisions, these include:

- *We need to develop processes to be able to handle cases of poor teaching*

  It should be noted that the original student request to release all CIOS comments to School Chairs is not being pursued. Rather, the Task Force focused the effort on the underlying need for accountability for poor teaching. The data, both on CIOS and on program-level data such as DFW rates, support that there is a problem with teaching among a small percentage of faculty. It is not acceptable to have a known problem and no processes in place to address that problem, especially when teaching is Georgia Tech's primary mission. The Task Force recommends that there be a process in place to investigate possible situations of poor teaching and to enact measures to ensure improvement to an acceptable level, including actions such as coaching and performance improvement plans with accountable outcomes. Student comments hold valuable information and add context to the numerical CIOS data. The goal of the investigation would be to pull together different sources of information, including student CIOS comments, to identify if there is a problem and what might be the extent and to help guide solutions if a problem exists. The resulting modified version of the original request on CIOS comments is much more targeted and limited in scope: allow administrators access to comments for the purpose of investigating problems with the course or teaching, and require those administrators to be trained on proper use of qualitative data that is possibly biased before having access to the comments. This modified request addresses most of the concerns expressed by faculty in the listening focus groups and town halls, and it addresses the underlying root cause for the original student request.

- *We need to have a process of teaching evaluations, when used in faculty evaluations, that is based on a solid rationale and is defendable. (CIOS, used as the sole measure, does not satisfy these conditions)*

  The main finding of the Task Force is that there is an overreliance on numerical CIOS scores as a measurement of teaching effectiveness. The CIOS represents opinions of students who may be biased and are not experts in either pedagogy or subject matter. For these reasons, the use of CIOS as the sole measure of teaching effectiveness is problematic in faculty evaluations since personnel decisions, such as annual reviews and promotion, should be made using processes that are defendable and are based on sound rationale. However, the CIOS does provide valuable information on the student experience in the class and is a good measure of *How We Teach*. The CIOS should be used in conjunction with at least one other metric that measures *What is Taught* or measures *Long Term Student Success* in order to span multiple dimensions of teaching effectiveness. One recommendation is to add an education statement to the promotion and tenure packages that contains self-evaluation similar to the one required for Georgia Tech Lecturers (in Appendix B). Many other universities are already taking the action to de-emphasize student evaluations and to combine them with other metrics such as self-evaluation and peer evaluation for a more balanced evaluation of teaching effectiveness. The American Sociological Association's positon statement on student evaluations of teaching (Appendix D) supports this finding and recommendation.

- *We need to have processes in place to mitigate bias or its effects in teaching evaluations*

  Another major concern for faculty has been bias in teaching evaluations. The focus was bias in CIOS, but bias can happen in many other forms of teaching evaluation including peer observation, peer evaluation of teaching materials, and student focus groups. Bias can occur in any subjective evaluation, such as review of applications for positions, review of archival publications, review of proposals, review of tenure and promotion cases, and review for awards. A university should pursue due diligence in mitigating bias or its effect in evaluations, and typically does so by requiring implicit bias training of evaluators, reviewing evaluation procedures, and reviewing outcome statistics for hiring cases. Some elements of due diligence should be done for the use of teaching evaluations as well, both in standard faculty evaluations and in determination of teaching awards. Note that the CIOS was evaluated for possible bias by examining the results for the "effective teacher" question and found that there was a slight but statistically significant difference between male and female instructors (0.08 out of 5). It is unlikely that this small difference would have any effect on promotion and tenure cases, but could be significant for teaching awards such as those determined by CTL and by the Faculty Honors Committee.

Additional recommendations outlined in the report include modifying the CIOS instrument as suggested in Appendix A to improve clarity, reduce redundancy, and add an inclusivity question.

Recommendations more suited to the school level include

- Encourage the use of mid-term evaluations to improve instruction during the semester
- Use more than one measure of teaching effectiveness when investigating possible problems with instruction or with a course.

- View teaching effectiveness on a broader scale than a particular instructor or course: to improve the quality and consistency of instruction across the department, utilize program-level metrics and peer evaluation (such as peer oversight of teaching materials or course coordination among peers).
- Create an alternative to CIOS for students to give feedback on broader topics to the department in a more timely basis than exit or alumni surveys, such as a web link to a feedback form.

For details of these recommendations, consult the body of the report. Also, please refer to Appendix E for a review of the literature on different ways of measuring teaching effectiveness.

References

These references are only those cited in the main section of this report. An extensive reference list is contained in the literature survey in Appendix E.

Cohen, P.A., (1981) "Student ratings of instruction and student achievement: A metaanalysis of multisection validity studies. Review of Educational Research, 51(3), 281–309, http://dx.doi.org/10.2307/1170209.

Doerer, K. (2019) "Colleges Are Getting Smarter About Student Evaluations. Here's How., Chronicle of Higher Education, *By* JANUARY 13, 2019 https://www.chronicle.com/article/Colleges-Are-Getting-Smarter/245457 Cohen, 1980; Feldman, 1989

Feldman, K. A. (1989). "The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies." Research in Higher Education, 30(6), 583–645.

Linse, A.R. (2017), "Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees," Studies in Educational Evaluation, Volume 54, September 2017, Pages 94-106, https://doi.org/10.1016/j.stueduc.2016.12.004

MacNell, L., Driscoll, A. & Hunt, A.N. (2015) "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching", Innov High Educ 40: 291. https://doi.org/10.1007/s10755-014-9313-4

Mitchell, K., & Martin, J. (2018). "Gender Bias in Student Evaluations." PS: Political Science& Politics, Cambridge University Press, 51(3), 648-652. https://doi.org/10.1017/S104909651800001X.

Peterson , D.A.M, Biederman, L.A., Andersen, D. , Ditonto, T.M. , Roe, K. (2019), "Mitigating gender bias in student evaluations of teaching," PLOT:ONE, Published: May 15, 2019, https://doi.org/10.1371/journal.pone.0216241

Uttl, B., White, C.A., Gonzalez, D.W. (2017) "Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related," Studies in Educational Evaluation, Volume 54, September 2017, Pages 22-42, https://www.sciencedirect.com/science/journal/0191491X/54/supp/C

Weinkle, L.J., Stratford, J.M., Lee, L.M.J. (2019) "Voice in Digital Education: The Impact of Instructor's Perceived Age and Gender on Student Learning and Evaluation", Wiley Online, Anatomical Sciences Education Research Report, first published: 28 January 2019 https://doi.org/10.1002/ase.1865

Appendix A: Recommended Changes to CIOS Questions

Changes to the current version are indicated by ~~deleted text~~ and new text

ITEM	SCALE
Student Effort	**Variable Scale**
1. On average, how many hours did you spend on this course per week (total in class, on homework, etc.)?	0 – 21+
2. What percentage of classes did you attend?	0 – 100
3. What percentage of the homework did you complete?	0 – 100
4. **Comments** about your responses in this section (eg., – were expected and expended effort appropriate for this course?). There is space for other overall comments later.	open-ended
Quality of Course	**5 Point Scale**
5. Rate how prepared you were to take this subject.	Completely Unprepared – Extremely Well Prepared
6. How much would you say you learned in this course?	Almost Nothing – An Exceptional Amount
~~7. Degree to which activities and assignments facilitated learning:~~	~~Very Poor – Exceptional~~
~~8.~~ 7. Degree to which exams, quizzes, homework (or other evaluated assignments) measured your knowledge and understanding:	Very Poor – Exceptional
~~9.~~ 8. Considering everything, this was an effective course.	Strongly Disagree – Strongly Agree
~~10~~ 9. ~~What was the best aspect?~~ What were the best features of the course, such as lectures, activities, assignments, and projects?	open-ended
~~11~~ 10. ~~How could it be improved?~~ How could this course be improved?	open-ended
~~12 **comments** about your specific responses in this section (quality of course). There is space for other overall comments later.~~	open-ended
~~13. **Overall Comments**~~	~~open-ended~~
Quality of Teaching	**5 Point Scale**
~~14.~~ 11. Instructor's clarity in discussing or presenting course material:	Very Poor – Exceptional

~~15.~~ 12. The instructor clearly communicated what it would take to succeed in this course.	Strongly Disagree – Strongly Agree
~~16.~~ 13. Instructor's respect and concern for students:	Very Poor – Exceptional
~~17.~~14. Instructor's level of enthusiasm about teaching the course:	Detached – Extremely Enthusiastic
~~18~~ 15. Instructor's ability to stimulate my interest in the subject matter:	Ruined My Interest – Made Me Eager to Learn More
~~19~~. 16. Instructor's availability for consultation:	Hard to Find – Highly Accessible
~~20.~~ 17. Helpfulness of feedback on assignments:	Not Helpful – Extremely Helpful
18. The instructor provided a positive environment in which to learn, regardless of my race, color, religion, sex, national origin, age, disability, sexual orientation, gender identity, or veteran status.	Strongly Disagree – Strongly Agree
~~21~~. 19. Considering everything, the instructor was an effective teacher.	Strongly Disagree – Strongly Agree
~~22.~~ 20. ~~What was the greatest strength?~~ What was the greatest strength of the instructor?	open-ended
~~23.~~ 21. ~~What is the most needed improvement?~~ How can instruction be improved?	open-ended
~~24. **Other comments** about your responses in this section (quality of teaching). There is space for other overall comments later.~~	open-ended

Appendix B: Format for the Georgia Tech Lecturer's Teaching Portfolio

The teaching portfolio used for the evaluation of Lecturers at Georgia Tech includes multiple measures of teaching effectiveness. Details of the components are listed below.

Reflection & Self-Evaluation

Give a narrative statement that describes your approach to teaching and other educational activities (such as new courses or course revision, co-curricular efforts, etc.). Describe your process for evaluating the effectiveness of your efforts and how you use that evaluation for improvement. For example, you may discuss how you improve your teaching using peer or CTL evaluations or using student comments from mid-semester course evaluations or CIOS. Frame your comments with specific examples from your teaching and/or educational activities. Include in your discussion, any innovations that you have tried, any teaching improvement activities (such as participating in teaching workshops), and how you keep your course material relevant and updated. As appropriate for your case, describe your motivation for making changes and the impact of those changes (on your teaching, on students, on your school or college, etc.). Describe your current strengths and areas for growth as well as how you create a culture of civility, collegiality, respect, and inclusiveness within the learning environments (classrooms, labs, studios, and/or supervised research).

Feedback & Evaluation from Others

1. Comments & letters from students: You may draw from your CIOS comments, Thank a Teacher notes, emails and letters from students, and so on.
2. Statement/documentation from peer review of teaching
3. Statement/documentation from peer review of teaching materials
4. Evidence of student learning (optional).Evidence might include things like pre- and post-test results, student performance indicators (e.g., rubric), percentage of students achieving learning outcomes in course, etc.

Course Surveys (CIOS) in the Institute table template

Sample Teaching Materials (Limit to 5 items) An appendix may be included that contains materials that explicitly support your narrative including, for example, items such as sample syllabi, class observation reports, graded student work, course assignments, rubrics, study guides, exams, etc.

Appendix C: CIOS Analysis and Pilot Survey

To evaluate the CIOS instrument, a series of correlative tests and exploratory factor analyses was conducted on the current CIOS scale items, looking for patterns in student responses. Results of this analysis showed that responses to current CIOS items tend to group into clusters with high inter-correlations, with a strong association to the overall effectiveness items. This suggested a fair amount of overlap, where items could be removed to make room for new items without significantly altering the general responses. Survey comments were also an important issue. While providing access to comments beyond the instructors was a key discussion, the current comment questions were thought to be vague in their instructions. This prompted two changes: A rewording of some existing prompts, and adding a comment that is available to program chairs.

In light of the importance of some CIOS elements to faculty review, proposed changes would need to be evaluated in terms of their possible effects on the ratings of current CIOS items. In order to test the impact of the alternate CIOS content, a pilot study was conducted during the Summer 2018 CIOS administration.

Procedure

To test the impact of the changed comment prompts, and the suggested added items, two CIOS Pilot instruments were developed. Both of the instruments included the comment label revisions, and the added instructor question on inclusivity. To look for a differential impact of adding an administrator-viewable item, this was added to only one of the Pilots (Pilot v1). The changes compared to the current CIOS are detailed in Table 1.

Table 1. Pilot Changes to CIOS

Standard CIOS (Control)	Pilot v1	Pilot v2
Instructor Items No changes	New Item: The instructor provided a positive environment in which to learn, regardless of my race, gender, sexual orientation, citizenship, disability, major, and other personal characteristics	New Item: The instructor provided a positive environment in which to learn, regardless of my race, gender, sexual orientation, citizenship, disability, major, and other personal characteristics
<u>Course Comments</u> What was the best aspect? How could it be improved?	What were the best features of the course, such as lectures, activities, assignments, and projects? How could this course be improved? New item: Is there anything that you would like to share with both	What were the best features of the course, such as lectures, activities, assignments, and projects? How could this course be improved?

	the instructor and academic administrators about this course?	
Instructor Comments What was the greatest strength? What is the most needed improvement?	What was the greatest strength of the instructor? How can instruction be improved?	What was the greatest strength of the instructor? How can instruction be improved?

An invitation to participate in the pilot study was extended to Summer 2018 CIOS course instructors, which netted 42 courses across four colleges. These courses were assigned to the two Pilot surveys by a stratified random assignment. In the process, Initial assignments were determined randomly, with later assignments placed to maintain a similar mix of course colleges and levels of instruction in each group.

For the control group, a subset of courses was selected from the Summer 2018 CIOS results. The courses were selected based on matching the pilot groups on several criteria: the colleges and departments represented in the study, and to the course levels represented in the pilot groups. In total, 153 courses were used for the Control group.

To increase the power available to detect differences, analysis was based on individual responses to the surveys, rather than using course composites. This yielded a total of 2,228 responses across the three Study Groups. The distribution of these responses is presented in Table 2.

Table 2. Study Groups

	Control	Pilot v1	Pilot v2	Total
College				
College of Design	128	1	6	135
College of Engineering	1005	217	147	1369
College of Liberal Arts	384	8	9	401
College of Sciences	102	153	68	323
Course Level				
1000	194	99	69	362
2000	377	85	22	484
3000	620	151	129	900
4000	299	26	9	334
6000	129	18	1	148
Class Size				
< 16	319	38	37	394
16-35	988	169	127	1284
36-99	276	155	39	470
> 99	36	17	27	80
Total	**1619**	**379**	**230**	**2228**

The primary concern in changing the instrument is the possibility that the altered question prompts and additional content might affect the scores of the current items. The "effective teacher" question is of particular concern as it is currently used for teaching evaluations in faculty Promotion and Tenure review, and the Class of 1940 Award. To be thorough, all course and instructor ratings were tested.

Because several factors have previously been found to associate with CIOS scores, an Analysis of Covariance (ANCOVA) was run for each CIOS item. In addition to the Study Group variable, College and Instructor Sex were added as additional independent variables, as well as course level and class size as covariates. To check for impact from the added administrator comment, two versions of the Study Group analysis was conducted: A three-level analysis (Control, Pilot v1, and Pilot v2) to look for any potential impacts, and a two-level analysis (Control vs. combined Pilots) to focus on the impact of the broader changes.

Results

For all items, there was no significant effect for the Study Groups in the three-level tests. For the two-level Study Group comparison a single Course item: *The degree to which activities and assignments facilitated learning*, $F(1,2168) = 3.929$, $p = .048$, partial $\eta^2 = 0.002$. Marginal means and *p*-values for the two-category analysis are presented below. No discernible differences were found on scale items between the two pilot groups.

Marginal Means and Between Subject significance by Group

ITEM	Means* Control	Means* Pilots**	*p*	partial eta squared
Course Questions				
13. How much would you say you learned in this course?	4.127	4.080	.839	.000
14. Degree to which activities and assignments facilitated learning:	4.237	3.964	*.048*	.002
15. Degree to which exams, quizzes, homework (or other evaluated assignments) measured your knowledge and understanding:	4.206	4.170	.735	.000
16. Considering everything, this was an effective course.	4.340	4.153	.215	.001
Instructor Questions				
4. Instructor's clarity in discussing or presenting course material:	4.402	4.304	.524	.000
5. The instructor clearly communicated what it would take to succeed in this course.	4.479	4.357	.728	.000
6. Instructor's respect and concern for students:	4.667	4.449	.194	.001
7. Instructor's level of enthusiasm about teaching the course:	4.735	4.552	.329	.000

8. Instructor's ability to stimulate my interest in the subject matter:	4.243	4.150	.556	.000
9. Instructor's availability for consultation:	4.506	4.346	.401	.000
10. Helpfulness of feedback on assignments:	4.278	3.996	.211	.001
11. Considering everything, the instructor was an effective teacher.	**4.469**	**4.395**	**.789**	**.000**

* Control factors: College. Covariates: Course level, Class size (actual enrollment)

** Combined Pilot groups

Appendix D: Statement on Student Evaluations of Teaching from the American Sociological Association

September, 2019

Downloaded 9/10/2019 from
https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_sept52019.pdf

Statement on Student Evaluations of Teaching

American Sociological Association
September 2019

Most faculty in North America are evaluated, in part, on their teaching effectiveness. This is typically measured with student evaluations of teaching (SETs), instruments that ask students to rate instructors on a series of mostly closed-ended items. Because these instruments are cheap, easy to implement, and provide a simple way to gather information, they are the most common method used to evaluate faculty teaching for hiring, tenure, promotion, contract renewal, and merit raises.

Despite the ubiquity of SETs, a growing body of evidence suggests that their use in personnel decisions is problematic. SETs are weakly related to other measures of teaching effectiveness and student learning (Boring, Ottoboni, and Stark 2016; Uttl, White, and Gonzalez 2017); they are used in statistically problematic ways (e.g., categorical measures are treated as interval, response rates are ignored, small differences are given undue weight, and distributions are not reported) (Boysen 2015; Stark and Freishtat 2014); and they can be influenced by course characteristics like time of day, subject, class size, and whether the course is required, all of which are unrelated to teaching effectiveness.

In addition, in both observational studies and experiments, SETs have been found to be biased against women and people of color (for recent reviews of the literature, see Basow and Martin 2012 and Spooren, Brockx, and Mortelmans 2015). For example, students rate women instructors lower than they rate men, even when they exhibit the same teaching behaviors (Boring, Ottoboni, and Stark 2016; MacNell, Driscol, and Hunt 2015), and students use stereotypically gendered language in how they evaluate their instructors (Mitchell and Martin 2018). The instrument design can also affect gender bias in evaluations; in an article in *American Sociological Review*, Rivera and Tilcsik (2019) find that the range of the rating scale (e.g., a 6-point scale versus a 10-point scale) can affect how women are evaluated relative to men in male-dominated fields. Further, Black and Asian faculty members are evaluated less positively than White faculty (Bavishi, Madera, and Hebl 2010; Reid 2010; Smith and Hawkins 2011), especially by students who are White men. Faculty ethnicity and gender also mediate how students rate instructor characteristics like leniency and warmth (Anderson and Smith 2005).

A scholarly consensus has emerged that using SETs as the primary measure of teaching effectiveness in faculty review processes can systematically disadvantage faculty from marginalized groups. This can be especially consequential for contingent faculty for whom a small difference in average scores can mean the difference between contract renewal and dismissal.

Given these limitations, the American Sociological Association, in collaboration with the scholarly societies listed below, encourages institutions to use evidence-based best practices for collecting and using student feedback about teaching (Barre 2015; Dennin et al. 2017; Linse 2017; Stark and Freishtat 2014). These include:

1. Questions on SETs should focus on student experiences, and the instruments should be framed as an opportunity for student feedback, rather than an opportunity for formal ratings of teaching effectiveness. For example, two universities – Augsburg University and University of North Carolina Asheville – recently revised and renamed their instruments to the "University Course Survey" and the "Student Feedback on Instruction Form," respectively, to emphasize that student feedback, while important, is not an evaluation of teaching effectiveness.

2. SETs should not be used as the only evidence of teaching effectiveness. Rather, when they are used, they should be part of a holistic assessment that includes peer observations, reviews of teaching materials, and instructor self-reflections. This holistic approach has been in wide use at teaching-focused institutions for many years and is becoming more common at research institutions as well. For example:
    - University of Oregon has undertaken a multi-year process to develop a holistic framework for assessing teaching effectiveness, including peer review, self-reflection, and student feedback. Extensive research and resources are available on the Office of the Provost website, including guidance on how to interpret SETs
    - University of Southern California has instituted peer review of teaching for faculty evaluation. Their Center for Excellence in Teaching provides resources for how to use peer review effectively and addresses common concerns.
    - University of California Irvine requires faculty to submit two types of evidence to document teaching effectiveness. In addition to SETs, faculty can submit a reflective teaching statement, peer evaluations of teaching, and other evidence like a Teaching Practices Inventory, developed by physicist Carl Weiman.
    - University of Nebraska Lincoln has articulated best practices for faculty evaluation that state, in part, "it is recommended that student evaluation scores should not be given undue weight in faculty evaluations, since these scores are easily manipulated and reflect many attitudes that extend beyond the successful accomplishment of the faculty member's teaching duties."
    - The University of Michigan's Center for Research on Teaching and Learning recommends that student ratings should never be used in isolation and should be part of a broader assessment of teaching effectiveness. They have developed resources that include a summary of research findings on SETs, a handout for students on how to make their feedback most helpful to instructors, and best practices for using SETs in personnel decisions.
    - Ryerson University has gone even further and is no longer using SETs for tenure or promotion decisions (Farr 2018). Instead, Ryerson asks faculty to compile a teaching dossier that includes a statement of teaching philosophy, evidence of curricular engagement, and self-reflections.
3. SETs should not be used to compare individual faculty members to each other or to a department average. As part of a holistic assessment, they can appropriately be used to document patterns in an instructor's feedback over time.
4. If quantitative scores are reported, they should include distributions, sample sizes, and response rates for each question on the instrument (Stark and Freishtat 2014). This provides an interpretative context for the scores (e.g., items with low response rates should be given little weight).
5. Evaluators (e.g., chairs, deans, hiring committees, tenure and promotion committees) should be trained in how to interpret and use SETs as part of a holistic assessment of teaching effectiveness (see Linse 2017 for specific guidance).

Gathering student feedback on their experiences in the classroom is an important part of student-centered teaching. This feedback can help instructors to refine their pedagogies and improve student learning in their courses. However, student feedback should not be used alone as a measure of teaching quality. If it is used in faculty evaluation processes, it should be considered as part of a holistic assessment of teaching effectiveness.

Endorsements

American Anthropological Association
American Dialect Society
American Folklore Society
American Historical Association
American Political Science Association
Association for Slavic, East European, and Eurasian Studies
Dance Studies Association
International Center of Medieval Art
Latin American Studies Association
Middle East Studies Association
National Communication Association
National Council on Public History
Rhetoric Society of America
Society for Cinema and Media Studies
Society for Personality and Social Psychology
Society of Architectural Historians
Sociologists for Women in Society

References

Anderson, Kristin J., and Gabriel Smith. 2005. "Students' Preconceptions of Professors: Benefits and Barriers According to Ethnicity and Gender." *Hispanic Journal of Behavioral Sciences* 27(2):184-20.

Barre, Elizabeth. 2015. "Student Ratings of Instruction: A Literature Review." Reflections on Teaching and Learning: The CTE Blog. Rice University Center for Teaching Excellence. Retrieved from https://cte.rice.edu/blogarchive/2015/02/01/studentratings.

Basow, Susan A., and Julie L. Martin. 2012. "Bias in Student Evaluations." Pp. 40-49 in *Effective Evaluation of Teaching: A Guide for Faculty and Administrators*, edited by Mary E. Kite. Washington, DC: Society for the Teaching of Psychology. Retrieved from http://teachpsych.org/ebooks/evals2012/index.php.

Bavishi, Anish, Juan M. Madera, and Michelle R. Hebl. 2010. "The Effect of Professor Ethnicity and Gender on Student Evaluations: Judged Before Met." *Journal of Diversity in Higher Education* 3(4):245-256.

Boring, Anne, Kellie Ottoboni, and Philip B. Stark. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *ScienceOpen Research* (DOI: 10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1).

Boysen, Guy A. 2015. "Significant Interpretation of Small Mean Differences in Student Evaluations of Teaching Despite Explicit Warning to Avoid Overinterpretation." *Scholarship of Teaching and Learning in Psychology* 1(2):150-162.

Dennin, Michael, Zachary D. Schultz, Andrew Feig, Noah Finkelstein, Andrea Follmer Greenhoot, Michel Hildreth, Adama K. Leibovich, James D. Martin, Mark B. Moldwin, Diane K. O'Dowd, Lynmarie A. Posey, Tobin L. Smith, and Emily R. Miller. 2017. "Aligning Practice to Policies: Changing the Culture to Recognize and Reward Teaching at Research Institutions." *CBE—Life Sciences Education* 16(5):1-8.

Farr, Moira. 2018. "Arbitration Decision on Student Evaluations of Teaching Applauded by Faculty." *University Affairs* August 28. Retrieved from https://www.universityaffairs.ca/news/news-article/arbitration-decision-on-student-evaluations-of-teaching-applauded-by-faculty/.

Linse, Angela R. 2017. "Interpreting and Using Student Ratings Data: Guidance for Faculty Serving as Administrators and on Evaluation Committees." *Students in Educational Evaluation* 54:94-106.

MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt. 2015. "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching." *Innovative Higher Education* 40(4):291-303.

Mitchell, Kristina M. W., and Jonathan Martin. 2018. "Gender Bias in Student Evaluations." *PS: Political Science and Politics* 51(3):648-652.

Reid, Landon D. 2010. "The Role of Perceived Race and Gender in the Evaluation of College Teaching on RateMyProfessors.com." *Journal*

of Diversity in Higher Education 3(3):137-152.

Rivera, Lauren A., and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84(2):248–274.

Smith, Bettye P., and Billy Hawkins. 2011. "Examining Student Evaluations of Black College Faculty: Does Race Matter?" *The Journal of Negro Education* 80(2):149-162.

Spooren, Pieter, Bert Brockx, and Dimitri Mortelmans. 2013. "On the Validity of Student Evaluation of Teaching: The State of the Art." *Review of Educational Research* 83(4):598–642.

Stark, Philip B., and Richard Freishtat. 2014. "An Evaluation of Course Evaluations." *ScienceOpen Research* (DOI: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1).

Uttl, Bob, Carmela A. White, and Daniela Wong Gonzalez. 2017. "Meta-Analysis of Faculty's Teaching Effectiveness: Student Evaluation of Teaching Ratings and Student Learning Are Not Related." *Studies in Educational Evaluation* 54(1):22-42.

Additional Resources

Association of American Universities. n.d. *Aligning Practice to Policies: Changing the Culture to Recognize and Reward Teaching at Research Universities.* Retrieved from https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/Aligning-Practice-To-Policies-Digital.pdf

American Educational Research Association. 2013. *Rethinking Faculty Evaluation: AERA Report and Recommendations on Evaluating Education Research, Scholarship, and Teaching in Postsecondary Education.* Retrieved from http://www.aera.net/Portals/38/docs/Education_Research_and_Research_Policy/RethinkingFacultyEval_R4.pdf.

Center for the Integration of Research, Teaching, and Learning (CIRTL). 2018. Local CIRTL Program Evaluation Resource Guide. Michigan State University. Retrieved from https://www.cirtl.net/resources/708.

Kite, Mary E., ed. 2012. *Effective Evaluation of Teaching: A Guide for Faculty and Administrators*. Washington, DC: Society for the Teaching of Psychology. Retrieved from http://teachpsych.org/ebooks/evals2012/index.php.

Stark, Philip B. 2018. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." Presentation given as part of the Teaching Assessment Working Group Speaker Series, Valuing Teaching: Challenges and Strategies to Value Effective Teaching. Simon Fraser University. Retrieved from https://www.youtube.com/watch?v=5haOjlfJDb8&feature=youtu.be.

Van Valey, Thomas L., ed. 2011. *Peer Review of Teaching: Lessons from and for Departments of Sociology*. Washington, DC: American Sociological Association.

Vasey, Craig, and Linda Carroll. 2016. "How Do We Evaluate Teaching? Findings from a Survey of Faculty Members." *Academe*, May-June. Retrieved from https://www.aaup.org/article/how-do-we-evaluate-teaching.

Appendix E: Literature Survey on Metrics and Measures of Teaching Effectiveness

Metrics and Measures of Teaching Effectiveness

Ruth Poproski and Rebekah Greene (2018)

Introduction

Measuring teaching effectiveness in higher education is an important but complex task. Over the last several decades there has been an increasing reliance on student evaluations of teaching (or SETs) as a method for evaluating the teaching role of instructors in higher education. At Georgia Tech this shift has been seen with the development of a focus on numerical results on the Course Instructor Opinion Survey (CIOS) at key points for review (e.g., promotion and tenure decisions, annual review, etc.).

More recently, the use of SETs as a measure of teaching effectiveness has been subject to significant scrutiny. Along with evidence-based concerns about gender bias in the results, it is clear that SETs are often difficult to interpret, and they focus on student perceptions that only sometimes align with teaching effectiveness. More importantly, it is clear that SETs are only one of several types of data that are required for a robust evaluation of teaching, and they are best used – for both summative and formative evaluation and for both personal development and career-related decision-making – when they are considered in the context of other measures of teaching effectiveness.

Berk (2018) argues that there are in fact fifteen different possible measures of teaching effectiveness, including student ratings, peer review/observation of teaching, peer review of course materials, teaching and course portfolios, employer/administrator/external expert ratings, advice from mentors, teaching awards, scholarship on teaching and learning, self-evaluation, learning outcomes measurement, exit and alumni interviews, and the use of videos. Selecting several of these measures rather than relying on just one can significantly improve the assessment of teaching effectiveness, and can lead to more robust understanding of ways for instructors to continue with their professional growth and development.

The purpose of this paper is to provide context for the use of multiple measures of teaching effectiveness, along with contextual framing and connection to the literature on teaching and learning. Specifically, we provide summary notes and references to specific studies to help the reader understand how and why each method might be used in the quest to evaluate teaching effectiveness.

Ruth Poproski is Assistant Director for Faculty Teaching and Learning Initiatives at the Center for Teaching and Learning. Rebekah Greene is a 2016-19 Marion L. Brittain Postdoctoral Fellow in the School of Literature, Media, and Communication. *This paper was written in conjunction with the work of Georgia Tech's 2017-18 Teaching Effectiveness Taskforce, and was supported by both the Center for Teaching and Learning and the Office of the Vice Provost of Graduate Education and Faculty Development.*

Table of Contents

Part I: Student Perspectives

Method 1: Student Focus Groups and Discussions

The focus group is a well-defined and regularly used research method that can be readily applied to teaching. In the evaluation of teaching, the *Small Group Instructional Diagnosis* (SGID) method has been developed as a best practice for gathering feedback from students about their experience in a course, and providing an opportunity for individuals to reflect on their teaching effectiveness. The act of working with a teaching and learning specialist to facilitate a focus group session, meeting with the facilitator to discuss the received student feedback, and communicating with students to address the feedback that they have provided can all aid instructors in adjusting and adapting their teaching approach to eliminate barriers to learning for their students. (Nelms 2015)

For more on the use of focus groups and SGIDs at Georgia Tech, visit http://ctl.gatech.edu/resources/best-practices/GnR/dialogue

SGIDs and focus groups are useful because they provide robust, targeted feedback from students about their ongoing experiences in a course. While there are similarities between end of semester student evaluations of teaching (SETs) and SGIDs, the timing and nature of the focus group approach allows for a closed feedback loop where instructors respond explicitly to student feedback, and adjust their teaching that can positively impact student learning throughout the course. In the process, instructors gain insight into student perceptions about their current teaching practices and are able to incorporate changes into their course structure as needed, benefitting current students. Ultimately, focus groups contribute to instructor development, providing an opportunity for the individual to reflect on their own teaching effectiveness, targeting current teaching strengths and opportunities for improvement identified by both students and teaching and learning specialists. (Black 1998; Coggman 1991; Diamond 2004).

Selected Studies: Using Focus Groups to Assess Teaching Effectiveness

The SGID method is a valuable technique for building effective teaching practices (Finelli et al 2008). This study highlights research conducted in the College of Engineering at the University of Michigan, investigating how different types of data (student ratings of teaching, student feedback collected during an SGID, or videotaped class sessions) can be utilized to assess gains in student ratings, faculty perceptions of consultations with instructional support staff, and reported changes in teaching practice. The researchers found that statistically the most effective consultation method was guided by student feedback from an SGID, seeing marked improvement in end-of-semester ratings related to “use of techniques to foster class participation” and “level of instructor respect towards students,” as well as “enthusiasm” (401) during the same semester that the SGID took place in.

Faculty participants reported adapting their approach and incorporating more effective teaching practices into their courses in several ways, including introducing more active learning, explaining concepts and/or examples more clearly, managing class time differently, calling directly on

students by name, and changing the pace of class. Key advantages associated with the use of SGIDs are that the faculty voice is represented in the process, gathered feedback is useful for instructors at all experience levels, and the use of SGIDs tends to increase teaching effectiveness over future semesters.

The SGID method is a beneficial formative assessment tool for faculty at all career stages (Blue & Evins 2008). This paper highlights the results of SGIDs administered in 45 courses to 1,086 students taught by 27 STEM instructors. The researchers reviewed data gathered from a selection of small (<15 students), medium (15-25 students), and large (>25 students) classes, and found that instructional strength and suggestion categories identified by students focused on teaching style/method, evaluation, class materials, instructor support/class policies, instructor characteristics, and organization. They note instructors can use the findings of SGID evaluations to improve current and future classes, particularly by clarifying policies with students, improving student ease of access to lecture notes and materials in preparation for exams, better managing class time in relation to complicated examples, and increasing patience with student questions. They also note that while instructors tend to request SGIDs before tenure and promotion decisions, the process has value for faculty at all career stages.

Using the SGID process to gather mid-semester feedback allows for adjustments to be made to a current course (Diamond 2004). Diamond (2004) reports results of a 2 year study of mid-semester feedback, suggests that feedback is a highly effective tool for improving teaching practices. The study consisted of 82 faculty from a wide range of disciplines, working with a facilitator to survey students. Facilitators guided student discussion and then consolidated results for faculty, which were then shared during follow-up sessions. In a follow-up survey, participating faculty (49 of the original 82) suggested that the program was valuable in helping them gain feedback which they then used to make changes to their classes. Participants reported changing in-class teaching techniques (35%), adjusting tests, assignments, and grading (31%), making changes to class content and how it's covered (16%), and clarifying details for students (10%). Ultimately, the majority of study participants reported finding the process useful, with many noting that it was helpful for them to be able to restructure their courses in such a way that current students could receive the maximum possible value from the course.

Method 2: Early and Mid-Course Feedback Surveys

Surveys are a quick and easy way to collect feedback from students – about their learning, their experience in the course so far, and their suggestions for changes to the way the course is taught. Unlike end of semester SETs, early and mid-course surveys are best used after about one-third of the semester has passed (typically around week four to six in a full-semester course), which allows enough time for students to form an opinion about a class, but also leaves enough time for the instructor to respond to the feedback gathered.

For more information about mid-semester feedback surveys, including several templates, visit http://ctl.gatech.edu/resources/best-practices/GnR/surveys

Selected Study: Using Early & Mid-Course Feedback Surveys to Assess Teaching Effectiveness

Mid-semester feedback surveys are a useful tool for assessing student learning and the effectiveness of teaching practices (Payette & Brown, 2018). Payette and Brown observe that mid-semester feedback is a systematic and formative mode of assessment that allows teachers to learn more about classroom dynamics, student engagement, and student experiences. Faculty can then use this data to consider adjustments to classes that are still in progress. While the typical process for mid-semester feedback is a collaborative effort between faculty and learning specialists, the authors note other variants (bare bones questionnaires, online surveys, or open feedback using Google Documents) are available for faculty who are short on time, or when critical staff are unavailable to partner with faculty. Finally, the authors note that mid-semester feedback yields benefits for faculty by revealing how their students are responding to course content and affording them the opportunity to make changes to content to facilitate learning.

Method 3: Longitudinal Evaluations – Alumni and Exit Interviews and Surveys

Surveys of exiting seniors and alumni are valuable tools for providing feedback on teaching. Although some variation in these surveys exists, with institutions choosing to design their own instruments or utilize pre-existing instruments such as the National Survey for Student Engagement (NSSE),[1] they typically provide useful information on curriculum design, teaching methods, evaluation techniques, and specific faculty strengths (Berk 2005). With more and more institutions utilizing alumni surveys or surveys of exiting seniors for accreditation purposes, such measures can also prove to be effective for the more regular assessment of teaching.[2] Encouraging alumni, especially recent graduates, to carefully reflect on the attributes and actions of teaching faculty – after they have gained some distance from the stresses of completing their degree – can provide a useful measurement of teaching effectiveness.

Selected Studies: Using Alumni and Exit Interviews and Surveys to Assess Teaching Effectiveness

Recent alumni can provide thoughtful insight into effective teaching practices (Moore & Kuol 2007). This article offers a consideration of how recent alumni can provide insight into effective teaching by identifying factors associated with excellent teaching – including communication practices and a focus on student needs. The researchers selected a sampling of 800 alumni and asked them to nominate teachers for awards and provide reasoning for these nominations. 139 responses were received, for a response rate of 17.4%. Alumni responses, focused on their experiences in the classroom, revealed the important contributions of teacher attributes and actions to the student learning process. Other response types included intellectual

[1] For more on the validity of the revised NSSE, see Zilvinskis, Masseria, & Pike, G.R. (2017).
[2] Dumford and Miller (2015) note a recent study (2014) by the National Institute of Learning Outcomes Assessment that indicates 85% of US higher education institutions use these surveys as part of their assessment mix, likely the outcome of mandates for increased engagement with alumni on the part of accreditation agencies such as ABET.

engagement and command over disciplinary areas, but this particular survey found that alumni especially valued an increased focus on students as a sign of effective teaching. As a result of this study, the authors recommend that properly constructed and implemented alumni surveys can be powerful indicators of what students value in their educational experiences.

Alumni surveys are useful for evaluating overall quality of instruction (Volkwein 2010). This article notes that feedback from alumni can provide valuable assistance building and developing performance and accountability systems for the evaluation of academic programs. Alumni feedback can be used for assessing curricula and support programs, and can also be useful for supporting accreditation and accountability efforts. Volkwein (2010) observes that typical alumni surveys may include questions focused on quality of instruction received, enhanced abilities and knowledge, satisfaction with academic experiences, perceived college impact on personal and professional development, and preparation for graduate school, career, or employment.

Using the answers to questions themed like this can aid faculty and administrators in strengthening teaching effectiveness. Volkwein (2010) makes several recommendations for the design of alumni surveys and sampling size, noting that surveys of recent graduates are the most valuable when considering teaching effectiveness efforts as "memories of the experience are fresh" (128). Volkwein (2010) also mentions several examples of well-designed alumni surveys, including those used by MIT, Penn State, and Georgia Tech.

Surveys of graduating students provide timely feedback regarding teaching effectiveness (Gainey 2017). This article suggests that surveys of exiting seniors can provide timely, relevant feedback that can help academic departments continuously improve. For example, responses can help department chairs "identify troubling trends and respond accordingly" (14), and they can help identify ways to "increase morale" among department members (16). Gainey (2017) recommends that surveys of graduating seniors should be short in order to encourage participation, and includes a sample survey distributed to graduating seniors in a management program. This survey asks students to rate the quality of the instruction they have received, and faculty availability. The survey also asks graduating seniors to identify the teacher who had the most positive impact on their studies. While the survey design does not allow for significantly developed or expansive responses, the answers to these questions do still provide an indication of how students perceive the effectiveness of the teaching they received as part of their major.

Part II: Peer Review of Teaching

As Yon, Burnap, & Kohut (2002) observe, "[t]he expanding use of peers in the evaluation of teaching is part of a larger trend in postsecondary education toward a most systematic assessment of classroom performance. Most scholars believe that informed and well-trained peers are ideally suited to evaluate their colleagues, especially colleagues in the same field" (104). Utilizing peer review can ultimately aid a program in building a more robust and systematic assessment program that does not solely rely upon student evaluations.[3] For example, by using

[3] As Fernandez & Yu (2007) note, students are good at observing basic daily classroom practice but peers

peer review of teaching through teaching observations, in combination with student evaluations and a peer review of course materials, faculty and administrators alike can specifically focus on developing and enhancing quality teaching practices. (Fernandez & Yu 2007; Yon, Burnap, & Kohut 2002)

See Table 1 for a breakdown of three main models for peer review as it is used in higher education.

Method 1: Peer Observation of Teaching

Faculty peer review of teaching – also known as peer observation, peer coaching, and peer feedback – has been a recognized tool for the measurement of teaching effectiveness in college classrooms for well over 35 years (Alabi & Weare 2014). The resulting significant body of literature suggests that the peer review of teaching can be utilized as part of a formative or a summative assessment process (Chism 2007). For example, peer observations of teaching can contribute to the overall evaluation of teaching effectiveness for the purpose of promotion and decisions (summative assessment), but they can also be used to aid instructors in their quest for professional growth and development, improving their teaching and classroom management techniques over time.

Ultimately, some peer observations of teaching will likely still be held by a program or department for summative assessment purposes having to do with personnel management issues such as tenure and promotion cases. However, recognizing and conducting peer reviews as a useful formative assessment tool centered on improving teaching and classroom management techniques can eventually increase faculty buy in to a more triangulated assessment process, one that is not solely based on student course evaluations.

For more information about the use of peer observations, visit http://ctl.gatech.edu/resources/best-practices/gnr/observation

Selected Studies: Using Peer Observations to Assess Teaching Effectiveness

Collaborative peer observations can improve teaching, foster intradepartmental collegiality, and increase collaboration among faculty (Fletcher 2018), This study was undertaken in an engineering department with approximately 45 faculty members, focused on developing and implementing a collaborative model for peer review. Study participants worked in pairs, and met in a pre-observation meeting to clarify how observations would be recorded and made available to the instructor. Each participant pair observed 4 total classes (2 per instructor) and then met together to provide feedback. Study participants ultimately used the feedback to improve their teaching, and cited an increased sense of collegiality within their department as a key benefit.

are more appropriate assessors of accurate course content and discipline appropriate teaching practices.

Table 1
3 models of peer review: evaluation, development, and collaboration (Fletcher 2018)

Feature	Evaluation Model	Development Model	Collaboration Model
Who is involved?	Senior staff, or evaluators/auditors selected to review other staff	Educational developers or expert teachers observe/review others	Teachers/peers/colleagues
Intent	Quality assurance; promotion, tenure, general review.	Demonstrate competency/improve teaching competencies	Improve teaching through dialogue; and reflection
Result	Report/judgment	Feedback /action plan for improvement to teaching	Analysis, reflection, discussion, teaching improvement
Relationship	Hierarchy of power	Hierarchy of expertise	Equality/mutuality
Confidentiality	Between manager and reviewee	Between reviewer and reviewee; may include manager	Between reviewer and the reviewee
Inclusion	Selected faculty: faculty being considered for tenure; applying for promotion; teaching award	Faculty on initial training course, faculty identified as needing teaching improvement	All involved in supporting student learning
Verdict	Pass/fail, score, quality assessment	Feedback on how to improve teaching	Non-judgmental, constructive & facilitated dialogue
Items reviewed	Teaching performance, course design, learning materials, student evaluations	Teaching performance, course design, learning materials	Any aspect of course design, teaching, student learning outcomes chosen by reviewee
Benefits	Institution, department	The reviewee (one way interaction)	Mutual benefits for both peers
Conditions for success	Effective management	Respected senior faculty	A culture in which teaching is valued and discussed
Hazards	Alienation, lack of cooperation, opposition	No shared ownership, lack of impact	Confirms existing practice, passive compliance

Peer observation of teaching is a valuable aid for reflection on one's teaching (Goldberg et al 2010). Researchers at Wichita State University surveyed 115 instructors in accredited communication science and disorders programs across the United States. The purpose of the survey was to investigate how these programs utilized the peer observation process, and for what purposes. While 27 instructors responded that peer observations were not currently required by their institutions, other respondents noted that peer observations are a regular part of their assessment mix. The majority of study respondents indicated that a follow-up discussion regarding their peer observation session was a pivotal part of the process, noting that it is this meeting and the resulting conversation about teaching practices that triggers modifications to teaching, rather than the actual observation itself. Respondents also reported that the act of conducting a peer review caused them to reflectively think about their own teaching.

Goldberg *et al* (2010) recommend the following best practices for peer observation of teaching:

- peer observations should be conducted by knowledgeable peers;
- the instructor being observed should be fully briefed regarding the process;
- the instructor being observed should be given time to prepare for the observation and should distribute class materials to the observer; and
- the instructor and observer should meet after the observation to collaboratively reflect upon and discuss suggested changes.

Peer observation of teaching is an inexpensive, flexible formative assessment tool for formative evaluation of teaching, and subsequent improvement of teaching effectiveness (Lowder et al 2017). Researchers at Kennesaw State University report in this article on a voluntary "Teaching Partners Program" offered between 2013-2014, involving 49 faculty pairs. Faculty members volunteered for the program, coordinated by the Center for the Excellence in Teaching and Learning. They were paired, typically with faculty from other departments, and then observed each other's classes and reported to each other on strengths or weaknesses that were observed. The observed benefits of this program included the following:

- inexpensive to develop and operate;
- faculty learn additional teaching strategies from each other;
- observing others provides younger faculty with confidence in their own teaching practices;
- increase in collegiality increases likelihood faculty continue to engage in peer observations;
- average time commitment needed on the part of an individual faculty member for such a process is less than five hours per semester;
- partnerships can be forged between faculty in the same department, similar disciplines, or across disciplines.

Method 2: Peer Review of Teaching Materials

Peer review of teaching materials typically involves review of artifacts like course syllabi, instructional plans, text selections, handouts, homework assignments, sample exams, tests, and projects, samples of graded student work, and framing comments from the instructor under

review. Ideally, the selected materials should represent how the instructor under review designs and implements their courses. The steps involved in this peer review process (e.g., preparing materials for review, receiving feedback, and engaging with the feedback to make changes) can provide instructors with valuable insight into their current teaching practices. Research also suggests that instructors who engage in this process tend to see an increase in their teaching effectiveness during subsequent semesters. (Berk 2005; Fink 2008; Houston 2010)

It is generally recommended that this process be used as a way to contribute to an instructor's professional growth and development – providing an opportunity for the individual to reflect on their own teaching effectiveness as they prepare for milestones related to the promotion and tenure process. While some recommend that the process of peer review of teaching materials be an annual event, others suggest that it be used instead for key check-in stages such as third-year review for junior faculty. Seasoned faculty can also benefit from the process as they are exposed to pedagogical innovations during the act of assessment.

Ultimately, the establishment of a department-wide peer review process brings about greater awareness of both teaching effectiveness and assessment practices, as materials are rated by peers on a scale—ideally developed and agreed upon department-wide prior to implementation-- which identifies instructor/material compliance with specific criteria (Berk 2005).

Selected Studies: Using Peer Review of Teaching Materials to Assess Teaching Effectiveness

The peer review of teaching materials process is useful for continued personal and professional faculty development (Thomas 2014). This article offers an examination of 34 studies of the peer review of teaching as read through a Strengths-Weaknesses-Opportunities-Threats (SWOT) framework. The researchers define the peer review of teaching as multi-pronged, including "the observation of lectures and tutorials, monitoring online teaching, examining curriculum design, and the use of student assessments" (113). They note that its strengths include that the technique "can be used as part of performance appraisal and tenure portfolios" and that it "sharpens individual skills" (113). Increased faculty participation in both the preparation of teaching materials and the assessment of these materials as part of a peer review process can aid individual continued growth in both teaching competence and cognitive understanding (115). Developing a peer review program including a teaching materials component can:

- help faculty gradually recognize the process as "useful for personal and professional development" (143);
- develop action plans for improvement based on both reflection and feedback from colleagues (148);
- "demonstrate professional responsibility and accountability" through the regular gathering of teaching materials for review purposes (150); and
- provide opportunities for faculty at all stages of their career to critically reflect on their own individual everyday teaching practice.

The peer review of teaching materials process develops good summative and formative assessment practices at the departmental level (Murphy & Flynn 2009). The authors of this study synthesize the research of Fink (2008) and Arreola (2007) – among others – suggesting that an effective and triangulated teaching quality assessment program can be readily built. Building on Arreola (2007)'s skill dimensions, defined as "content expertise, instructional design skills, instructional delivery skills, instructional assessment skills, and course management skills," Murphy & Flynn (2009) design a portfolio system for the assessment of teaching materials at their institution. They propose this as a way to supplement the use of SETs and peer observations in the evaluation of teaching effectiveness. A post-study survey of participants noted that some identified the process as a challenging one, primarily due to issues relating to how precisely to evaluate materials. The researchers note that the majority of identified challenges can be overcome by:

- selecting specific criteria for assessment (preferably via an "open discussion process" at the departmental level;
- ensuring that all departmental reviewers are clear as to how to evaluate the materials as related to the identified criteria; and
- encouraging a wider campus discussion (among both faculty and administrators) about best practices regarding critical reflective practice as an important part of the overall assessment process.

Involving all faculty in the process at the department level provides an increased attention to the quality of teaching. Junior faculty maintain their progress towards tenure and promotion while senior faculty can stay abreast of new and evolving developments in pedagogy.

Preparing for a review of teaching materials fosters continuous reflection and growth in effective teaching practices (Fink 2008). In this chapter, Dee Fink considers the evaluation of teaching and how it can be used as a form of summative and formative assessment. Fink outlines how teaching materials and their careful evaluation by peers can supplement student evaluations. Fink (2008) also includes a candid discussion of how the preparation of teaching materials can foster a continuous cycle of formative assessment (and improved teaching) in the individual faculty member, especially if materials are centered around the following four general criteria: knowledge of subject matter, design of learning experiences, student interaction, and course management.

Part III: Instructor Reflection and Review

Method 1: Teaching Portfolios

A teaching portfolio is a factual description of an instructor's teaching strengths and accomplishments. It includes documents and materials that collectively suggest the scope and quality of a professor's teaching performance. The portfolio is to teaching what lists of publications, grants, and honors are to research and scholarship. When portfolios are submitted for personnel decisions, the focus should be on evidence that documents the professor's best

works as a teacher and demonstrates that significant student learning (cognitive or affective) has taken place. The faculty member's achievements, awards, and successes are the focus.

Teaching portfolios can also contribute to instructor development as they provide an opportunity for the individual to gather together data and artifacts related to their teaching, and reflect on their own teaching effectiveness. In addition, assembly of a teaching portfolio can help instructors identify gaps or weaknesses in their approach to teaching, potentially prompting them to pursue new opportunities for professional development, and/or to try new things with their teaching. (Seldin, Miller, & Seldin 2010)

Selected Studies: Using Teaching Portfolios to Assess Teaching Effectiveness

Teaching portfolios have two goals: development and evaluation (De Rijdt 2006). This study examines faculty attitudes towards teaching portfolios, and suggests that the portfolio ideally has two goals: development and evaluation. A study of teaching portfolios at four different institutions raised a variety of concerns including faculty knowledge of portfolio design, the time required to complete a portfolio, and the transparency of the assessment process. Respondents generally agreed that teaching portfolios are a useful way to reflect on and improve teaching and that there are some positive consequences of keeping a portfolio, including more clarity about responsibilities and course construction. Ultimately the study found that those faculty respondents using portfolios as a form of assessment tool improved their teaching via reflection, improving course materials, and the feedback of reviewers.

Teaching portfolios are valuable instruments to assess thinking about teaching (Trevitt & Stocks 2012). The authors of this study utilize their experiences in aiding others in developing teaching portfolios and in assessment to argue that portfolios aid assessors in gaining insight into the faculty member's day-to-day pedagogy, which they term as "authenticity." They provide a detailed rubric based on their experiences, suggesting that assessors value portfolios that contain markers of authenticity such as treatments of teaching contexts/issues and challenges faced in the classroom, evidence of the development, and practical use of teaching approaches, evaluations and instructor response to feedback.

The researchers use a small study of portfolios to demonstrate how these criteria for authenticity can be measured against three different levels of performance, ultimately suggesting that teaching portfolios can be so varied in terms of their content that assessment teams will need to come to a consensus on a methodology for assessment. Their research suggests that reflective portfolios are valuable instruments to "assess thinking about teaching" due to "the narrative element [which] is generally more revelatory than the various bits of paperwork that lecturers use to evidence their practice".

Assessing teaching portfolios requires development of specific procedures (Tigelaar et al 2005). This study considers the portfolio as a collection of qualitative data that is sometimes hard to objectively rate. The researchers use a review of literature relating to teaching portfolios and their assessment to make recommendations for criteria that can be used to effectively judge the rich variety of items on display in many faculty portfolios, ultimately arguing that candidates under

review should be made fully aware of the assessment process. They also offer the following recommendations for organizational procedures for assessment:

- portfolios should be read by at least two assessors who represent multiple perspectives on teaching;
- assessors should keep a record of the evaluation process, use guiding questions, and discussion with each other during the process;
- mentors should offer advice about the choice of items for inclusion;
- internal and external monitoring should provide checks and balances; and
- purpose of the portfolio should be explained.

While the researchers do recognize that all of these components may be hard to implement, they note that doing so provides the candidate with deep feedback that will aid in their continued development and representation of effective teaching practices, and it can allow the portfolio to serve as a measure for both formative and summative assessment.

Method 2: Self-Evaluation of Teaching Effectiveness

Over the past two decades, more and more higher education institutions have been considering faculty self-ratings (sometimes also called self-evaluations or self-assessment) as evidence of teaching effectiveness. These self-ratings, frequently taking the form of annual progress reports, record teaching accomplishments over the course of an academic year. Instruments for self-ratings may include structured forms that document the type of course taught, number of students, display teaching objectives, activities, accomplishments, and failures. Checklists and writing prompts for reflection may also be used as a part of the process.

Examples & Samples

- Case Studies: Arreola (2007)
- Sample Self-Reflection Tool: Lorch (2013)
- Sample Annual Report Structure: Hoyt & Pallet (1999)

The literature generally agrees that self-ratings should be utilized with other measures of teaching effectiveness, such as portfolios or peer observations of a class. In sum, the work of organizing materials for self-ratings on an annual basis and critically reflecting upon teaching efforts can lead to overall increases in effectiveness and aid faculty in documenting their individual achievements in preparation for retention, tenure, and promotion review. (AAUP 2015; Berk 2005; Buskist 2006; Miller & Seldin 2014; Seldin 1982; Seldin & Miller 2009).

Selected Studies: Use of Self-Evaluation to Assess Teaching Effectiveness

Self-assessment instrument preparation leads to increases in teaching effectiveness through reflection (Rigler et al 2016). This article reports the results of adding a faculty self-assessment[4] instrument into a merit evaluation system at a state university. A two-year study

[4] The researchers rightly point out that the body of current literature is "unclear on the distinction between self-evaluation, self-assessment, and self-reflection" (3), with authors regularly using all three terms interchangeably. Rigler and his colleagues observe some distinctions, suggesting that self-evaluation offers faculty insight into teaching that other sources cannot provide and shapes instruction. They also state that

involving 19 faculty members in a single department revealed that, for the majority of participants, utilizing a PDF form allowed faculty to clearly identify accomplishments, set goals for the following year, and continue reflecting on opportunities for individual growth.

The researchers note that self-assessment instruments can add value to evaluation procedures as they allow faculty to:

- identify their achievements in teaching;
- utilize their reflections on teaching practices to increase awareness of the effect of their practices and to design future experiments in the classroom; and
- consider ways to advance teaching effectiveness in the classroom.

Self-evaluations provide valuable insight into teaching (Miller & Seldin 2014). This article offers an overview of Miller and Seldin's 2010 survey of academic deans and the assessment of faculty performance. The researchers surveyed 538 academic deans of accredited liberal arts colleges and received 420 responses to their questionnaire. Miller & Seldin (2014) observe that the surveyed deans value classroom performance as the most important part of faculty performance, but also note that research, publication, campus committee work, and student advising are also highly valued. They state that self-evaluation of teaching has increased from 2000 (58.7%) to 2010 (67.6%) and suggest that "self-evaluation can provide insights into the values and beliefs that help shape course and instructional objectives and, in turn, contribute to classroom competency."

Self-reflection helps faculty members think about effective teaching and future development (Lorch 2013). This selection of self-reflection tools developed for occupational therapy teaching faculty in cooperation with a learning consultant offers an overview of basic core competencies for effective teaching. Of special note is their faculty self-reflection tool, that can be readily adapted for any discipline. The self-reflection tool asks faculty to consider what they most enjoy teaching, and also encourages a self-identification of teaching challenges.

Method 3: Self-Evaluation using Videos of Instructor's Teaching

Recording, watching, and reflecting upon videos of one's own teaching can all be useful techniques for increasing teaching effectiveness. Videos are useful measures of teaching effectiveness – whether they are a three minute clip of a lesson or a full class period in length – because they provide documented evidence of a faculty member's command of a classroom. Videos are relatively easy to create and aid instructors in understanding how their classroom personas may impact student learning. Watching a video of a class and then reflecting on speaking rates and tone, volume, body language, or usage of classroom technology can help improve teaching effectiveness through increased attention to classroom management of student

self-assessment allows faculty to "rate their effectiveness on a scaled form or provide brief written evaluations of their teaching performance" (Centra 1979 qtd. in Rigler *et al* 2016, p. 4) and is more formal in nature. Finally, self-reflection considers teaching practices, student results, and potential implications for future teaching (Bullock and Hawk as qtd. in Rigler *et al* 2016, p. 4). Given that this particular study refers to the testing of a specific, formal process, self-assessment is the preferred term for the researchers.

understanding of concepts. In the long run, developing careful criteria for how the video should be produced and how it will be assessed (either at the departmental or institutional level) will assist faculty in more readily adapting this easy, cost-effective method that offers constructive suggestions for continued improvement. (Berk 2005; Snoeyink 2010; Tripp & Rich 2012a).

Selected Studies: Using Videos of Teaching to Assess Teaching Effectiveness

Using videos as a measure of teaching effectiveness can help teachers make modifications to classes (Snoeyink 2010). This study of pre-service teachers suggests video as a self-analysis tool can be a valuable aid for increasing teaching effectiveness. Citing literature on the usage of videos in teaching education programs, the author notes that video can be viewed multiple times for clarity, can help teachers view themselves from different viewpoints, and can aid teachers in becoming more mindful during the act of teaching and "modifying their actions in the moment" (102). Additional benefits include becoming more aware of body language, facial expressions, posture, usage of whiteboards and other classroom technologies, rates of speaking, tone of voice and volume, and repetition of words.

Study participants reported growth in three important areas (classroom management, student understanding of concepts, and self-reflection) after viewing their videos. The author rightfully observes that the relatively small size of the sample (8 participants) make the results difficult to measure and interpret but suggests that there is still value in utilizing video as a self-analysis tool for teachers of all levels.

Videos have value as documented evidence of teaching performance (Berk 2005). Berk (2005) claims that the use of video can be a powerful, relatively easy to create tool for improving teaching effectiveness. Faculty can readily record their performance in the classroom, using colleagues, institute videographers, audiovisual experts, or even a tripod. The recording should take place in the regular classroom and should be as natural as possible for best results. Berk (2005) proposes multiple options for video evaluators, including self and/or a peer or group of colleagues who view the video and provide ratings and feedback. On the value of this process, Berk notes that it is a powerful documentary of teaching performance, a good way to generate a profile of positive and negative behaviors, and it leads to the development of specific objectives to address deficiencies in the classroom.

Videos are useful tools for reflection on teaching (Tripp & Rich 2012b). Tripp & Rich (2012b) reviewed 63 studies where participants recorded their own teaching, examined their performance on video, and reflected on their performance. They note that there are a number of factors that need to be carefully considered when designing a form of video assessment, including facilitation of the reflection process, the extent to which instructors reflect on their video individually or collaboratively, and the number of reflections required (see Table 2 for more details). They also note that there is no current consensus about what length of video is best for use as a reflection tool – from a simple 3 minute video to a full class session length. Further, they suggest that guiding reflection through the use of written rubrics or collaborative reflection with other peers can be important, especially for junior faculty.

They conclude that "video-aided teacher reflection has demonstrated posited change through varied measures (self-report, case studies, lesson plans, pre-/posttest scores)" (679), but also note that we lack frameworks for use of videos as a measure of teaching effectiveness (688).[5]

Table 2
Summary of the dimensions of the video analysis process, and recommendations for future research (Tripp & Rich 2012b)

Dimension	Definition	Question
Reflection tasks	Tasks teachers participated in during or after viewing their teaching: (1) completing codes or checklists, (2) participating in interviews or conferences, (3) writing reflections and (4) video editing.	What type of reflection tasks will I ask teachers to engage in during their video reflections
Guiding reflection	How the reflection process was facilitated. For example, in some studies participants chose their own reflection focus, while in other studies researchers or supervisors guided the teachers' reflections.	Will I provide teachers with a framework to guide their reflections?
Individual/ collaborative reflection	Individual reflection refers to instances where teachers viewed and reflect on their video individually.	Will I ask teachers to reflect individually, collaboratively, or both?
Video length	In past studies the length of video used for reflection varied from 3 minutes to an entire teaching episode.	What length of video will teachers use for reflection?
Number of reflections	In past studies the number of times teachers reflected on their videos varied from one to more than three reflections.	How many times will teachers reflect on their videos?
Measuring reflection	This refers to how studies determined the influence of video on teachers' reflections.	What methods will I use to determine if video was beneficial for teacher reflection?

[5] For more on how Tripp and Rich have used video to improve teaching at their home institution, see Tripp & Rich (2012a).

Method 4: Scholarship and Reflection

One way for instructors to engage in formative evaluation of their own teaching is to apply a research mindset to the enterprise, and to follow that up with a critically reflective consideration of how their observations can and should impact their teaching practices. Shulman (2000) calls this our professional obligation, and points out that scholarship on teaching and learning is necessary to ensure that one's work as an educator continues to improve. As he notes, "Active scholarship of teaching provides the teacher with a very different perspective on what he or she may have been doing for many years." (48). As Williams (2015) puts it, "good teachers use research (or variations of it) to inform their teaching every day, sometimes without knowing it. They are constantly reviewing assessment data, making critical observations of their students and themselves, and collecting qualitative and quantitative data." (4). Brookfield (1995) argues that the path to discovering the worth of your teaching is through a process of critical reflection, connected to observable data. In other words, reflective scholarship on teaching and learning can be a powerful way to evaluate (and improve upon) one's own teaching effectiveness.

According to Trigwell (2013), scholarship on teaching and learning is frequently described in the literature as a way or means to:

- enhance university teaching;
- raise the status of teaching;
- come to teach more knowledgeably;
- assess the quality of teaching;
- enhance students' experience of learning;
- enhance the research profile of an individual or department; and
- stimulate interest in teaching.

Selected Study: Using Scholarship and Reflection to Assess Teaching Effectiveness

Approaching teaching with scholarly inquiry makes teaching more effective (Trigwell 2013). Trigwell (2013) reports on a questionnaire distributed to 56 faculty at a major Australian university regarding the scholarship of teaching and learning (SOTL). Participants in the study were surveyed on six dimensions, including use of literature, articulation of a teaching/learning model, inquiry, reflection, teaching as a public activity, and peer review. Respondents who indicated that they use a student-focused approach to teaching as opposed to an information-transfer based approach scored higher on all of the dimensions and also were found to be more reflective about their teaching practices overall. While the author suggests that further research into the use of SOTL for assessment purposes is needed, the results of the questionnaire indicate that approaching teaching with scholarly inquiry can make both teaching and student learning more effective.

Part IV: Evidence of Student Learning (Learning Outcomes)

Identifying and evaluating what student learning has occurred – also known as measuring student learning outcomes – is a valuable way to measure teaching effectiveness because it gets at the heart of the purpose of teaching. Measuring learning outcomes can help us understand the

knowledge, skills, abilities, and habits that a student has gained at the end of a project, course, program, or major. As a result, the use of learning outcomes is considered one of the most pivotal measures of teaching effectiveness for both summative and formative assessment. (Suskie 2018)

There are multiple methods for assessing learning outcomes, including (but not limited to) the following:

- questionnaires administered at the beginning and end of a class or capstone experience (a.k.a. pre- and post-tests);
- focus groups coordinated by faculty and/or teaching center personnel;
- the use of portfolios with instructor reflections; and
- faculty submission of student materials that map to specific outcomes (e.g., quizzes, exams, presentation films, etc.).

Engaging with learning outcomes in a formative assessment process can directly benefit the individual instructor, but in order for the maximum benefit to be derived from this process, they need to be intimately involved in the process of determining which direct or indirect methods will be used to evaluate the effectiveness of their teaching by way of learning outcomes. Learning outcomes assessment can also help instructors intentionally develop their approach to teaching in ways that will increase student learning. (Bresciani 2009; Cydis 2015; Duque & Weeks 2010; Palmer 1998)

At both the departmental level and the broader institutional level, evidence about student abilities is needed for quality assurance purposes, especially when considering effective teaching practices in a summative assessment process (Tam 2014). Many accreditation organizations, such as ABET or AACSB, have rigorous standards in place to ensure that students are learning the skills and attitudes that they need to be successful in the professional workforce. Other stakeholders, including the University System of Georgia's Board of Regents, also value learning outcomes as a measurement of teaching effectiveness. As Kuh et al (2014) notes, “[c]learly articulated learning outcomes are important in determining whether students know and can do what an institution promises and what employers and policy makers expect.” (p. 8).

Selected Studies: Using Learning Outcomes to Measure Teaching Effectiveness

Using learning outcomes builds consistency and clarity in the classroom, increasing teaching effectiveness and easing the assessment process (Tam 2014). This article offers an overview of how learning outcomes can be used for quality assessment and curriculum improvement in higher education. Tam (2014) offers the following three levels at which assessment might take place:

- *student* (knowledge, skills, and abilities gained through engagement in a teaching/learning experience);
- *program/course* (the development or growth as a result of studying a particular course or program/attainment of competencies); and
- *institute* (“attainment against established standards (criterion-referenced assessment) or as the performance of an individual or group compared to others (norm-referenced assessment)” (161).

All three levels of outcomes contribute to the enhanced development of a student-centered learning experience. Aligning learning outcomes with teaching and learning activities not only ensures consistency within and across the curriculum, but can also foster teaching effectiveness through the continued act of assessment.

Tam (2014) also observes that there are four main benefits to an outcomes-based approach:

- clarity (clear communication between stakeholders about what kind of learning is expected at the end of a course or program);
- flexibility (multiple methods or modes of delivery can be deployed in the classroom; different direct and indirect measures of learning can be assessed;
- comparison (established standards across programs or institution used for accreditation, benchmarking, and accountability);
- and portability (making it easier for students to transfer credits between institutions).

While Tam (2014) notes that it can be hard to measure certain types of outcomes, she concludes that outcomes-based learning is a valuable measure of learning effectiveness and instructional quality, helping instructors succeed in improving both student learning and their own pedagogical practice.

Learning outcomes measurement should not be used for summative evaluation of teacher effectiveness – but may have value for formative and programmatic assessment (Berk 2014). Berk (2014) discusses the use of learning outcomes as a measure of teaching effectiveness, and concludes that while this approach can be used for the formative evaluation of teaching, it is a mistake to use this approach for the summative evaluation of teaching and applying it to personnel decisions (promotion, tenure, annual review, etc.). He identifies three options for developing tools to measure learning outcomes, and explains why none of these options is suitable for the evaluation of an individual instructor's teaching (see Table 3).

Further, Berk (2014) points out that even when we measure learning outcomes by computing (for example) gain scores across multiple measures, the interpretation of those results does not necessary correlate with teaching effectiveness. That is, in many cases student performance is not related to teaching performance, and so the use of learning outcomes to measure teaching effectiveness is misplaced.

Learning outcomes are easily used to guide changes at the course level, enhancing teaching effectiveness (Kuh et al 2014)*.* This report from the National Institute for Learning Outcomes Assessment (NILOA) addresses how higher education institutions across the US are using student learning outcomes as part of their assessment process. A survey was administered to 1,202 provosts (or their designates) during the spring and summer of 2013, with 43% responding to questions about current assessment activities and the usage of evidence directly relating to student learning outcomes. Kuh *et al* (2014) notes that there has been a steady increase in the assessment of learning outcomes over the past decade. Survey participants report the following perspectives, based on their experiences at their institutions:

- more direct faculty involvement in the assessment process is necessary as it helps improve teaching and learning;
- leveraging and sharing what people from around the institution are doing in their

Table 3
Approaches to measuring learning outcomes

Measure	Problem
Instructor-designed measures (e.g., in-class assessments)	Not typically tested for reliability and validity (and the process of doing so is too onerous for widespread practice).
Perceived learning measures (e.g,. self-assessments by students with respect to their perceived learning)	Results tend to correlate with student achievement (e.g., grades), but not with actual learning.
Standardized Tests	It is difficult to find well-designed tests that also align with the curriculum and instruction in particular courses, and for many disciplines and courses there are no such tests available.

classrooms can improve teaching and learning;

- faculty and staff may need clarity when it comes to defining both learning outcomes and assessment methods used to measure outcomes;
- outcomes-based assessment can provide evidence for the need for other improvement-related tasks related to increased teaching effectiveness, such as curriculum modification and faculty development;
- outcomes-based assessment results can be used to guide changes at the course level, which, in turn, increases teaching effectiveness;
- faculty should be encouraged to design assignments that address student learning within the course and simultaneously provide evidence that can be used for institutional purposes; and
- outcomes can be measured effectively using classroom-based assessment, national student surveys, or rubrics.

Kuh *et al* (2014) also notes that the survey's respondents recognize the issue of faculty buy-in to the assessment process, including concerns regarding the validity of assessment measures, how assessment results may or may not be used for performance evaluations, and how to use data for to inform a path toward improvement. The researchers suggest that the majority of these concerns can be readily addressed, noting the following conclusions:

- increased faculty involvement in the process should be encouraged and recognized;
- institutions should work on building assessment of authentic learning into the everyday work of faculty, utilizing processes that may already be in place; and
- institutions should use collected data more frequently to work with faculty to develop and improve teaching effectiveness.

Through better cooperation between institutional stakeholders, a continuous culture of enhanced attention to student learning may be achieved.

Part V: Conclusion

As noted at the beginning of this paper, there are a wide variety of methods that can be used to evaluate the effectiveness of teaching. Much has been written about the use of SETs as a measure of teaching effectiveness, and opinions vary as to their placement in formal reviews of individual instructors. There is general consensus that SETs should not be used in isolation, and that multiple measures should be used in order to effectively evaluate teaching. Further, there is general agreement that while the student voice can be captured through the use of end-of-semester SETs, there are additional ways to capture that voice, and it is also important to capture data from instructors, their peers, and general evidence of student learning.

In addition to the various methods discussed above, some recommend the use of teaching awards as a measure of teaching effectiveness (cf. Berk 2018). This, however, is an under-studied approach to measuring teaching effectiveness, and many have raised concerns about its use. For more information about the use of teaching awards as a measure of effective teaching, please see the appendix to this paper.

Finally, it is worth noting that the assessment of teaching effectiveness need not occur in isolation. Individual instructors can turn to faculty mentors and specialists from the Center for Teaching and Learning. Similarly, School Chairs and others tasked with making personnel decisions and contributing to the promotion and tenure system are also able to turn to experienced colleagues and the Center for Teaching and Learning for support. When making curricular and programmatic decisions, the Office of Academic Effectiveness can also provide support, bringing expertise in assessment to the table.

As Evans & Bertani Tress (2009) remind us, many faculty place a high value on the teaching component of their work, and typically desire to excel in all areas of their jobs. One of the great things about Georgia Tech is the rigor and passion with which all stakeholders attack their work. When it comes to assessing teaching effectiveness – whether it be for individual gain or more formal purposes – there are multiple avenues through which we can engage with this same rigor and passion. By use of a robust plan, we can lead the way to effective assessment of teaching.

References

AAUP. (2015). Statement on teaching evaluation. In *Policy Documents and Reports*. 11th ed. Baltimore: Johns Hopkins UP, 219-22.

Alabi, J. and Weare, W.H. (2014). Peer review of teaching: Best practices for a non-programmatic approach. *Communications in Information Literacy*, 8(2): 180-191

Arreola, R.A. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems*. (3rd ed.) New York, NY: Wiley.

Berk, R.A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17 (1), 48-62.

Berk, R.A. (2014). Should student outcomes be used to evaluate teaching? *Journal of Faculty Development*, 28 (2), 87-96.

Berk, R.A. (2018). Start spreading the news: use multiple sources of evidence to evaluate teaching. *Journal of Faculty Development, 32*(1), 73-81.

Black, B. (1998). Using the SGID method for a variety of purposes. *To Improve the Academy*, 17, 245-262.

Bledsoe, R.S. and Richardson, D.S. (2016). An innovative award catches faculty in the act of great teaching. *Journal on Centers for Teaching and Learning*, 8, 51-69.

Blue, J., Wentzell, G.W., and Evins, M.J. (2008). What do students want? Small group instructional diagnoses of STEM faculty. *Proceedings of the 2014 Physics Education Research Conference. Englehardt, P.V., Churukian, A.D., and Jones, D.L., Editors.* 43-46. Retrieved from doi:10.1110/perc.2014.pr.007

Bresciani, M. (2009). Evaluating the efficiency, effectiveness and sustainability of outcomes-based program review. *Journal of Faculty Development*, 23(1), 30-39.

Brookfield, S. (1995). Becoming a Critically Reflective Teacher. San Francisco: Jossey-Bass.

Buskist, W., et al. (2006, April 1). Evaluating and improving your teaching. *Observer*. Retrieved from https://www.psychologicalscience.org/observer/evaluating-and-improving-your-teaching

Carusetta, E. (2001). Evaluating teaching through teaching awards. *New Directions for Teaching and Learning*, 88*,* 30-40.

Chism, N. (2006). Teaching awards: What do they award?. *The Journal of Higher Education*, 77 (4), 589-617.

Chism, Nancy (2007). *Peer Review of Teaching, A Sourcebook*. 2nd ed. Anker Publishing: Bolton (MA).

Coggman, S.J. (1991). Improving your teaching through small-group diagnosis. *College Teaching*, 39 (2), 80-82.

Cydis, S., et al. (2015). Integrating and assessing essential learning outcomes: fostering faculty development and student engagement. *Journal of the Scholarship of Teaching and Learning*, 15(3), 33-52.

De Rijdt, C., et al. (2006). Teaching portfolios in higher education and their effects: an explorative study. *Teaching and Teacher Education*, 22, 1084-1093.

Diamond, M.R. (2004). The usefulness of structured mid-term feedback as a catalyst for change in higher education classes. *Active Learning in Higher Education*, 5 (3), 217-231.

Dumford, A.D., and Miller, A.L. (2015). Are those rose-colored glasses you are wearing? Student and alumni responses. *Research and Practice in Assessment* 10, 5-14.

Duque, L.C. and Weeks, J.R. (2010). Towards a model and methodology for assessing student learning outcomes and satisfaction. *Quality Assurance in Education*, 18(2),84-105.

Evans, L. and Bertani Tress, M. (2009). What drives research-focused university academics to want to teach effectively? Examining achievement, self-efficacy, and self-esteem. *International Journal for the Scholarship of Teaching and Learning*, 3 (2) 1-17.

Fernandez, C.E. and Yu, J. (2007). Peer review of teaching. *Journal of Chiropractic Education*, 21 (2), 154-161.

Fletcher, J.A. (2018). Peer observation of teaching: a practical tool in higher education. *Journal of Faculty Development* 32 (1), 1-14.

Finelli, C., et al. (2008). Utilizing instructional consultations to enhance the teaching performance of engineering faculty. *Journal of Engineering Education*, 97 (4), 397-411.

Fink, L.D. (2008). Evaluating teaching: a new approach to an old problem. *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development*. Vol. 26. San Francisco: Jossey-Bass, 3-21.

Gainey, T.W. (2017). Improving quality in academic units: the important role of senior surveys. *Quality Approaches in Education*, 8 (1), 12-20.

Goldberg, L.R., et al. (2010). Peer review: the importance of education for best practice. *Journal of College Teaching and Learning*, 7 (2), 71-84.

Hammer, D., et al. (2010). Recognition of Teaching Excellence (Article 164). *American Journal of Pharmaceutical Education*, 74 (9), 1-11.

Houston, N.M. (2010). Organizing your teaching materials. *Chronicle of Higher Education*. Retrieved from http://chronicle.com June 2018.

Hoyt, D. P., and Pallett, W. H. (1999). Appraising teaching effectiveness: Beyond student ratings (IDEA Paper #36). Manhattan, KS: The IDEA Center. Retrieved from www.theideacenter.org/sites/default/files/Idea_Paper_36.pdf June 2018.

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing What Students Know and Can Do: The Current State of Student Learning Outcomes Assessment in US Colleges and Universities*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Lorch, A. (2013). The teaching self-reflection tool and skills checklist. *Department of Occupational Therapy Faculty Papers*. Paper 59. Retrieved from http://jdc.jefferson.edu/otfp/59/ June 2018.

Lowder, L., et al. (2017). Peer observation: improvement of teaching effectiveness through class participation at a polytechnic university.*Journal of STEM Education*, 18 (4), 51-56.

Malfroy, J. and Willis, K. (2018). The role of institutional learning and teaching grants in developing academic capacity to engage successfully in the scholarship of teaching and learning. I*nternational Journal for Academic Development*, 1-12.

Marchant, T. and Wallace, M. (2016). Gender balance in teaching awards: evidence from 18 years of national data. *Journal of Higher Education Policy and Management*, 38 (4), 393-405.

Miller, E. J., and P. Seldin. (2014). Changing practices in faculty evaluation. *Academe* 100 (3): 35–38.

Moore, S. and Kuol, N. (2007). Retrospective insights on teaching: exploring teaching excellence through the eyes of the alumni. *Journal of Further and Higher Education*, 31 (2), 133-143.

Murphy, T., MacLaren, I., and Flynn, S. (2009). Toward a summative system for the assessment of teaching quality in higher education. *International Journal of Teaching and Learning*, 20 (2), 226-236.

Nelms, G. (2015). Small Group Instructional Diagnosis. Chapter 5 in *Assessing The Teaching of Writing*, edited by Amy E. Dayton. USA: University Press of Colorado.

Palmer, P. (1997). The heart of a teacher: identity and integrity in teaching. *Change*, 29 (6), 15-21.

Payette, P.R., and Brown, M.K. (2018). Gathering mid-semester feedback: three variations to improve instruction. IDEA Paper #67. Manhattan (KS): Center for Faculty Evaluation and Development.

Rigler, K., Cook-Benjamin, L., Wieland, R., and Tholstrup, C. (2016). The inclusion of self-assessment in merit evaluation. *Cogent Education* (3), 1-12.

Shulman, L. (2000). From Minsk to Pinsk: Why a scholarship of teaching and learning? *The Journal of the Scholarship of Teaching and Learning*. 1: 48-52.

Seldin, P. (1982) Self-assessment of college teaching. *Improving College and University Teaching*, 30 (2), 70-74.

Seldin, P., and Miller, J.E. (2009). The academic portfolio: A practical guide to documenting teaching, research, and service. San Francisco: Jossey-Bass, 31-51.

Seldin, Peter J., Elizabeth Miller, and Clement A. Seldin (2010). *The Teaching Portfolio: A Practical Guide to Improved Performance and Promotion/Tenure Decisions*. San Francisco, CA: Jossey-Bass.

Snoeyink, R. (2010). Using video self-analysis to improve the "withitness" of student teachers. *Journal of Computing in Teacher Education*, 26 (3), 101-110.

Suskie, L. (2018). *Assessing Student Learning: A Common Sense Guide*.(3rd ed.) San Francisco (CA): Jossey-Bass.

Svinicki, M. and Menges, R. (1996). Consistency within diversity: guidelines for programs to honor exemplary teaching. *New Directions for Teaching and Learning*, 69, 109-113.

Tam, M. (2014) Outcomes-based approach to quality assessment and curriculum improvement in higher education, *Quality Assurance in Education*, 22 (2), 158-168.

Thomas, S., et. al. (2014). A qualitative review of literature on peer review of teaching in higher education: an application of the SWOT framework. *Review of Educational Research*, 84 (1), 112-159.

Tigelaar, D., Diana H.J.M. Dolmans, Ineke H.A.P. Wolfhagen, and Cees P.M. van der Vleuten (2005). Quality issues in judging portfolios: implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education*, 30(5), 595-610.

Trevitt, C. & Stocks, C. (2012). Signifying authenticity in academic practice: a framework for better understanding and harnessing portfolio assessment. *Assessment & Evaluation in Higher Education*, 37(2), 245-257.

Trigwell, K. (2013). Evidence of the impact of scholarship of teaching and learning purposes. *Teaching and Learning Inquiry*, 1 (1) 95-105.

Tripp, T. and Rich, P. (2012a) The influence of video analysis on the process of teaching change. *Teaching and Teacher Education*, 28, 728-739.

Tripp, T., and Rich, P. (2012b). Using video to analyze one's own teaching. *British Journal of Educational Technology*, 43 (4), 678-704.

Volkwein, J.F. (2010). Assessing Alumni Outcomes. *New Directions for Institutional Research*, 125-139.

Williams, K.W. (2015). *Doing Research to Improve Teaching and Learning: A Guide for College and University Faculty*. NY, NY: Routledge.

Yon, M., Burnap, C., and Kohut, G. (2002). Evidence of effective teaching: perceptions of peer reviewers. *College Teaching, 50*(3), 104-110.

Zilvinskis, J., Masseria, A.A., and Pike, G.R. (2017). Student engagement and student learning: Examining the convergent and discriminant validity of the revised national survey of student engagement. *Research in Higher Education*, 58 (8), 880-903.

Appendix: Teaching Awards as a Measure of Teaching Effectiveness

Many institutions of higher learning are now recognizing quality teaching through an awards process. Chism (2006) states that the growth in teaching awards, especially over the past four decades, can be traced to (1) many institutions' wish to acknowledge their support for teaching, (2) the importance of acknowledging the accomplishments of excellent teachers, and (3) encouraging other faculty to strive for excellence in their own teaching. Depending on the type of award, data is typically collected from three primary sources: past and present students, peers, and the individual nominee.

Faculty participation in teaching award processes through the establishment of award criteria, participation in portfolio development workshops, and the adjudication of awards tends to increase faculty awareness of effective teaching practices, and reflection on their own teaching. In addition, the process of updating materials is good practice for the type of self-reflection valued in the formative assessment process and can ultimately lead to increases in teaching effectiveness. (Bledsoe & Richardson 2016; Carusetta 2001; Marchant & Wallace 2016; Svinick & Menges 1996).

That said, an individual's receipt of teaching awards is not generally recommended as a good measure of teaching effectiveness. For example, these awards are often highly politicized, they are small in number (so many very effective teachers will never receive a teaching award), and the criteria associated with the awards often do not align with direct measures of teaching effectiveness. In addition, in recent years there has been a growing concern that the percentage of awards won by female faculty and by part-time/contingent faculty is not appropriately representative. As a result, the use of teaching awards in the evaluation of teaching effectiveness – particularly for summative purposes – is not well-placed. (Chism 2006, Marchant & Wallace 2016).

Selected Studies: Using Teaching Awards to Assess Teaching Effectiveness

Teaching awards come in multiple forms and can advance both scholarship and teaching effectiveness (Malfroy & Willis 2018). This article suggests that teaching awards can come in multiple forms, including institutional learning and teaching grants. Malfroy & Willis (2018) suggests that while such grants do have value as they lead to innovations in teaching, the development of resources (including collaboration), and increased frameworks for good practice, research on evaluating the effectiveness of such grants has been somewhat limited to date. For grant programs to be truly effective, clearly identified expected outcomes need to be established early in the process. They studied 27 completed grant funded projects from 2013 and 2015 and learned that final reports detailing the results of funded projects varied widely in terms of literature

support prior to updated guidelines. The new guidelines for final reports ask grant recipients "to refer to relevant literature but to also explain how the project advanced the existing knowledge and practice," resulting in more substantial information (7). Study participants ended in sharing the results of their grant-funded projects at conferences, in book chapters, and in journal articles (among other venues), which suggests that this form of teaching award can help increase effectiveness by providing opportunities to share the results of research that happens inside the classroom.

Teaching awards and required evidence: Steps to improving teaching excellence (Hammer 2010). This white paper from the 2008-2009 *American Association of Colleges of Pharmacy Task Force for the Recognition of Teaching Excellence* considers best practices for acknowledging and rewarding teaching excellence. The task force members note that effective teaching recognition programs must clearly identify the purpose, criteria, number and mix of awards, frequency, type of award, and method of nominating and determining awardees when establishing programs. They establish useful guidelines for using these types of evidence and also include a list of useful questions to ask when establishing awards and determining the form the award itself should take. While they suggest that the connection between teaching awards and a renewed commitment to teaching improvement is currently critically under-researched, they also acknowledge that other opportunities associated with the teaching award process do have value in improving overall teaching effectiveness.

The teaching award process contributes to collegial responsibility and uses data from multiple sources (Svinicki & Menges 1996). This article offers a list of suggestions for ensuring that teaching award programs not only recognize effective, excellent teaching, but also remain above suspicions of favoritism. These suggestions include:

- establishing consistency with the institution's mission and values and communicating those values to the community;
- grounding award programs in research-based teaching competencies rather than having them be dependent on special interests, favoritism, or popularity;
- recognizing all significant facets of instructional activities that are conducted by faculty (including laboratories, fieldwork, office hours, course design, etc.);
- rewarding collaborative as well as individual achievements;
- neither precluding nor displacing rewards for teaching that are part of the institutionalized reward system;
- calling on those who have been honored to continue to contribute to the development of others through peer modeling;
- contributing to collegial responsibility for promoting exemplary teaching through the act of asking faculty to nominate colleagues;
- encouraging self-reflection at all levels of the institution;
- basing the program on sound assessment practices, including multiple data sources, multiple measures, and consistency over time; and
- keeping the program open to scrutiny and change as conditions evolve.