Student Course Evaluation Improvement Project

PRESENTED BY	**Thomas Cohen, Chair**
REVIEW DATES	**SEC – April 8, 2019	SENATE – April 24, 2019 & October 2, 2019**
VOTING METHOD	**In a single vote**
RELEVANT POLICY/DOCUMENT	**N/A**
NECESSARY APPROVALS	**Senate, President**

ISSUE

In January 2017, the Associate Provost of Learning Initiatives and Executive Director of the Teaching & Learning Transformation Center (TLTC) submitted a proposal to the Senate Executive Committee (SEC) recommending improvements to the CourseEvalUM system. The proposal noted that the University's method of evaluating courses had not changed significantly since an electronic system was first introduced in 2008 and called for a comprehensive review of the existing approach to evaluating courses and instructors. In February 2017, the SEC voted to charge the Academic Procedures & Standards (APAS) Committee with reviewing past Senate action on course evaluations, reviewing scholarship on course assessments and practices at Big 10 and peer institutions, consulting with a range of stakeholders across campus, and recommending whether changes should be made to the current system.

RECOMMENDATIONS

The APAS Committee makes a series of recommendations regarding changes to the University's CourseEvalUM system as shown in the attached report.

COMMITTEE WORK

In spring 2017, the APAS Committee met with the proposers to discuss their concerns with course evaluations and researched past Senate action on course evaluations. It also met with representatives of the Office of Institutional Research, Planning, and Assessment (IRPA), which oversees implementation of course evaluations, and the Course Evaluation Advisory Group, which advises IRPA on the development of the system and its uses on campus. A subcommittee with members from the APAS Committee and members from the Course Evaluation Advisory Group was formed to consider the charge in-depth and make recommendations to the APAS Committee. The subcommittee met approximately twenty times between July 2017 and January 2019, and consulted with a range of subject-matter experts and stakeholders.

The subcommittee developed recommendations based on its review and submitted its report to the full APAS Committee in February 2019. APAS shared its preliminary directions with the stakeholders the subcommittee previously consulted, as well as with the Senate at its March 2019

meeting. It also consulted with the Office of General Counsel. The committee developed a final set of recommendations based on the feedback it gathered. After due consideration, the APAS Committee voted to approve its recommendations on the course evaluation system at its meeting on March 29, 2019.

ALTERNATIVES

The University could continue using the current CourseEvalUM system.

RISKS

There are no known risks.

FINANCIAL IMPLICATIONS

There may be some cost associated with implementation of the recommendations. Specifically, there could be costs associated with incorporating survey results into tools used by students when registering for classes depending on how the recommendation is implemented.

Student Course Evaluation Improvement Project

2018-2019 Committee

Thomas Cohen (Chair)
Deanna Barath (Graduate Student)
Progyan Basu (Faculty)
Lauren Brown (Undergraduate Student)
William Cohen (Ex-Officio Provost's Rep)
Linda Coleman (Faculty)
Adrian Cornelius (Ex-Officio University Registrar)
Jeffrey Franke (Ex-Officio Graduate School Rep)
Lee Friedman (Faculty)
Shannon Gundy (Ex-Officio Rep for Director of Undergraduate Admissions)
Agisilaos Iliadis (Faculty)
Lisa Kiely (Ex-Officio Undergraduate Studies Rep)

Byung-Eun Kim (Faculty)
Roberto Korzeniewicz (Faculty)
Marilee Lindemann (Faculty)
Celina McDonald (Faculty)
Benjamin Parrish (Undergraduate Student)
Julian Savelski (Exempt Staff)
Michael Sparrow (Exempt Staff)
David Straney (Faculty)
Elizabeth Warner (Faculty)

Date of Submission

April 2019

BACKGROUND

In January 2017, the Associate Provost of Learning Initiatives and Executive Director of the Teaching & Learning Transformation Center (TLTC) submitted a proposal to the Senate Executive Committee (SEC) recommending improvements to the CourseEvalUM system. The proposal noted that the University's method of evaluating courses had not changed significantly since an electronic system was first introduced in 2008, and pointed out several areas where the current CourseEvalUM system could be improved. It called for a comprehensive review of the existing approach to evaluating courses and instructors, one that would be informed by recent scholarly literature and intended to revise the course evaluation items used. In February 2017, the SEC voted to charge the Academic Procedures & Standards (APAS) Committee with reviewing past Senate action on course evaluations, reviewing scholarship on course assessments and practices at Big 10 and peer institutions, consulting with a range of stakeholders across campus, and recommending whether changes should be made to the current system (Appendix 9).

CURRENT PRACTICE

In 2002, the Senate considered a proposal relating to teaching evaluations, which led to the formation of a Joint Provost/Senate Task Force on Course Evaluations and Teaching in spring 2003. The task force presented a preliminary report and recommendations in February 2004; the Senate subsequently passed a resolution recommending that "there be a University-wide requirement for student evaluations in all undergraduate and graduate courses." The task force submitted its final report in April 2005, identifying four purposes for a new course evaluation system:

- Formative Evaluation: To provide diagnostic feedback to faculty for the improvement of teaching.

- Summative Evaluation: To provide one measure of teaching effectiveness for use in the Appointment, Promotion, and Tenure (APT) and post-tenure review processes and in annual productivity reviews.

- Informative Evaluation: To provide information to students for their use in the selection of courses and instructors.

- Outcome Evaluation: For the purposes of documenting student learning.

The task force also recommended steps to enhance the institution's ability to assess and improve curriculum and instruction. The Senate voted to approve the Recommendations for the Implementation of Web-based Student Course Evaluations (Senate Document #02-03-39) in December 2005. The Provost subsequently appointed an implementation committee, which presented items for a new University-wide course evaluation system to the Senate as an informational item in April 2006. A complete overview of past Senate action on course evaluations can be found in Appendix 1. Today, practices and priorities associated with the CourseEvalUM system are informed by the Course Evaluation Advisory Group, a body composed of representatives from each College/School and various administrative units.

There are currently sixteen CourseEvalUM survey items about instruction that are asked of students in all courses, and four that are asked about teaching assistants (Appendix 2). Fifteen of the items about instruction are forced-choice items, and one is open-ended; three of the teaching assistant items are forced-choice, and one is open-ended. Items focus on either the course or the instructor. In cases where there are multiple instructors, the instructor-specific items are asked for each individual. In addition, Colleges/Schools and departments have the ability to add additional items to the evaluations; those that do so are included in Appendix 3. In some cases, the number of these additional items is significant and more than doubles the length of the instrument. Additional items are most often used to gather insights on courses and instructors and align with the original purposes behind course evaluations; in some instances, however, they are used to gather data for accreditation purposes, and may have little or nothing to do with the course being evaluated.

Results from different items on the evaluations are available to different groups. Responses to eight of the items (Administrator Items) are only visible to instructors and authorized campus administrators, and are intended for use in evaluating and improving instructor performance. The Administrator Items also include a single open-ended item. Given that responses to the Administrator Items can inform personnel decisions, they are kept confidential and only made available to the instructor and relevant administrators. In order to ensure that students benefit from the system, eight additional items are included (Student Items) that are only visible to students and instructors. This separation is known as the "firewall." Results from these items are primarily intended to help students select courses. There is some overlap in the themes addressed by the Administrator and Student Items.

Course evaluations are administered near the end of each term, and conclude before the start of the final exam period. For standard, fifteen-week courses, the system opens two weeks before the last day of classes. Results are not available until after final grades have been submitted. Reports to instructors and administrators include the score distribution, average, and standard deviation for each item. Additionally, comparative averages by College/School, department, and course level are reported. An "overall score" summarizes the average of all five Likert-scale Administrator Items.

Presently, results dating back to 2007 are available to currently registered students. In 2014, however, the University adopted a new vended platform to conduct evaluations. The ability for students to view results gathered after 2014 was not implemented until fall 2018, meaning students have only recently been able to access results from courses offered in the last several years. The current platform cannot show student grade distributions, which were previously available. Results

for courses with five or fewer students are not made available to students, and students can only view results for a particular course/section if the response rate exceeds 70%. Over the past three years, University response rates in fall and spring semesters have ranged from about 55% to 60%, thereby making results from many courses inaccessible to students. Additional information on how CourseEvalUM results are used may be found in the Committee Findings section.

COMMITTEE WORK

In spring 2017, the APAS Committee met with the proposers to discuss their concerns with course evaluations and researched past Senate action on course evaluations. It also met with representatives of the Office of Institutional Research, Planning, and Assessment (IRPA), which oversees implementation of course evaluations, and the Course Evaluation Advisory Group. A subcommittee with members from both the APAS Committee and the Course Evaluation Advisory Group was formed to consider the charge in-depth and make recommendations to the APAS Committee. Subcommittee members included:

Phil Evers (faculty, APAS Chair from 2016-2018)

Susan Hendricks (faculty, past APAS member)

Marilee Lindemann (faculty, APAS member)

Michael Passarella George (staff, Assistant Director for Decision Support in IRPA/Course Evaluation Advisory Group member)

Doug Roberts (faculty, past APAS member)

Joseph Sullivan (faculty, Course Evaluation Advisory Group member)

Kaci Thompson (staff, Course Evaluation Advisory Group member)

The subcommittee met approximately twenty times between July 2017 and January 2019, and consulted with a range of subject-matter experts and stakeholders. The subcommittee:

- Reviewed past Senate action establishing the purposes of course evaluations;

- Reviewed research on course evaluations at UMD conducted by IRPA, including research on what our current items measure, bias in course evaluations, and how results are used by students in the course selection process;

- Met with experts in the field, including Dr. Sandra Loughlin, an educational psychologist who directed the Office of Transformational Learning in the Robert H. Smith School of Business; and Dr. Alice Donlan, an educational psychologist and Director of Research for the TLTC;

- Reviewed a survey of current literature on student assessments of teaching;

- Met with the Associate Provost for Faculty Affairs and the Council of Associate Deans for Faculty Affairs (CADFA) to discuss the use of course evaluations in making teaching assignments and improving instructor effectiveness;

- Reviewed the TLTC's new Mid-Semester Evaluation of College Teaching (MSECT) pilot;

- Consulted with both undergraduate and graduate students at two meetings of the Senate Student Affairs Committee, as well as a focus group of students;

- Evaluated possible replacement items through cognitive interviews with students conducted by Loughlin and Donlan; and

- Reviewed practices at Big 10 and peer institutions.

The subcommittee developed recommendations based on its review and submitted its report to the full APAS Committee in February 2019. In reviewing the recommendations, APAS considered whether or not the practice of conducting University-wide assessments of teaching effectiveness should be discontinued altogether. Ultimately, the committee determined that collecting University-wide data on student experiences serves a useful purpose that should be continued and improved. Its recommendations are premised on a belief that the University should take this opportunity to reimagine the instrument in light of current best practices; narrow its purpose; and develop a tool that provides more and better information to administrators, instructors, and students.

APAS shared its preliminary directions with the stakeholders the subcommittee previously consulted, as well as with the Senate at its March 2019 meeting. It also consulted with the Office of General Counsel, and sent a survey to a select group of department chairs to gather feedback on its proposed recommendations for making teaching assignments and evaluating instructor performance. The committee considered the subcommittee's recommendations and the feedback it gathered in order to develop a final set of recommendations. After due consideration, the APAS Committee voted to approve its recommendations on the course evaluation system at its meeting on March 29, 2019.

PEER INSTITUTION AND SCHOLARLY RESEARCH

Big 10 and Peer Institutions
The committee reviewed information provided by the proposer on course evaluation practices at Big 10 and peer institutions (Appendix 4). Most peer institutions have a set of campus-wide questions that are asked, and many allow colleges, departments, and sometimes instructors to include additional questions. Several articles published over the course of the committee's work indicate that institutions have identified concerns and are conducting reviews similar to the ones called for in APAS' charge (see Flaherty, 2018; Supiano, 2018; Doerer, 2019). The committee found that while institutions have adopted a range of approaches, it is difficult to identify any single instrument or set of best practices the University might adopt. Well-designed instruments are developed for particular contexts and to answer specific questions. In consultation with campus experts, the committee determined that the University should first identify what it wants to assess regarding courses and instructors, and then design or adapt an instrument that targets those elements as precisely as possible.

Scholarly and Institutional Research
The committee reviewed recent literature relevant to student evaluations of teaching, including studies addressing bias in teaching evaluations. While documenting bias can be difficult, the literature suggests that broad or vague items, and items that allow open-ended comments in particular, are more susceptible to bias (Felton et al., 2008; Lindahl and Unger, 2010; Porter, 2011). Studies also indicate that student learning is not highly correlated with student perceptions of teaching (Uttl et al., 2017). The committee found that much of the literature on student evaluations is particular to the instrument being studied, and does not necessarily yield broadly applicable insights (Linse, 2017). The committee's work was also informed by a memo from Dr. Sandra Loughlin reviewing literature on student evaluations of teaching (Appendix 5). The memo noted that such evaluations often ask students about things for which they are not the best source of data.

The committee also reviewed several studies IRPA has conducted on the current course evaluation system that investigate usability, reliability and validity of the items; the relationship between response rates and instructor scores; and whether an instructor's race/ethnicity/citizenship and gender explain differences in ratings. IRPA also conducted phone interviews to identify practices associated with higher response rates, finding that students are more likely to complete evaluations when instructors emphasize that they value the feedback and are interested in improving a course. These studies are summarized in Appendix 6.

COMMITTEE FINDINGS

Administrator & Instructor Use of Course Evaluation Results

In the course of its review, the committee consulted various administrators, including the Associate Provost for Faculty Affairs and the Senior Vice President and Provost. It also attended several meetings of CADFA and the Undergraduate Academic Programs Committee (UGAP). Once it had identified preliminary recommendations, the committee distributed them, along with a survey, to a select group of chairs from both large and small departments, as well as the dean of a non-departmentalized College.

The committee learned that course evaluation results are used by administrators and instructors to assess performance in tenure and promotion cases. They are also used to identify potential problems in an instructor's teaching and indicate where additional intervention may be necessary. Practices vary, sometimes significantly, when it comes to the way results are used in appointment and promotion processes, though the committee identified the following generalities.

- **Tenured/tenure-track (T/TT) faculty:** The role that course evaluation results play in the APT process varies by College/School, though the University has been shifting to more holistic evaluations of faculty teaching that involve peer evaluations, student mentoring, curriculum/course development, a teaching portfolio, and other instructional activities. The Office of Faculty Affairs provides a template that can be used when compiling and interpreting the numeric results of evaluations; responses to open-ended items are typically included at the discretion of the individual faculty member. In general, however, the committee found that course evaluation results play a limited and decreasing role in the APT process, particularly given teaching is only one aspect on which a T/TT faculty member is assessed.

- **Professional track (PTK) faculty:** The role course evaluations play in the Appointment, Evaluation, and Promotion (AEP) process for PTK faculty is more significant than it is in APT. There is no consistent standard for peer evaluation process for PTK faculty, and course evaluations are, in some cases, the sole or most significant factor used in making renewal or promotion decisions, particularly for purely instructional faculty.

Most administrators are aware of the results' limitations and contextualize them with other sources of data on instructor performance. This caution is not universal, however, and evaluation results are sometimes used in inappropriate ways. These include averaging all of an instructor's results into a single number for comparisons with peers or to give teaching awards, and comparing instructor averages to the College/School- and department-averages for courses of a similar level.

Instructors and administrators often use the open-ended comments to contextualize and nuance the numeric data; as one administrator put it: "the numbers tell you there is a problem and the comments tell you what the problem is." Some instructors expressed significant concerns with the

comments, however, noting that they can be biased and even abusive. Some instructors also indicated that they ignore the comments entirely. Both IRPA and the Office of Faculty Affairs have received complaints from instructors about the open-ended comments. To help address these concerns, IRPA has added the following language before the open-ended items on the evaluations:

> *In order to help instructors and administrators best use your feedback to improve teaching and learning at the university, please be thoughtful and constructive when writing comments. Inappropriate or offensive comments do not reflect the civil engagement we value as an institution, and they are generally not effective in stimulating improvements to instruction. Should any comments take the form of actual threats, they will be forwarded, with the student's identifying information, to campus police for threat assessment.*

While it has received requests from instructors to remove specific comments, IRPA is not equipped to evaluate the nature of student comments or make decisions regarding which comments should be removed. In addition, the current system does not offer a way to easily delete comments. The committee acknowledged that the use of open-ended comments can be abused, but determined that their value was significant enough that they should be retained. The committee recommended including two open-ended items that use specific prompts related to positive aspects and areas for improvement. The committee hopes that this will yield more actionable responses and may reduce the number of biased comments.

Student Use of Course Evaluation Results
The committee met twice with the Student Affairs Committee and hosted a small focus group of students from different disciplines. From these sessions, the subcommittee gained insights into how students use course evaluation results and other resources to select courses. Students reported using a range of resources including CourseEvalUM results, third-party websites (among them ratemyprofessor.com, ourumd.com, and planetterp.com), and word of mouth. Student responses to a 2016 Campus Assessment Working Group Snapshot indicated 43% of students considered CourseEvalUM a "major factor" when choosing courses (Appendix 7). Students expressed uncertainty as to whether the results were for instructors, administrators, or other students, and did not always understand which items referred to the course and which to the instructor. The distinction between Administrator and Student Items was also unfamiliar.

When asked what would make a course evaluation system more useful and improve completion rates, students asked for access to the open-ended comments, grade distributions, and a "star system" for providing a simple snapshot of student satisfaction with a course or instructor. Students reported that incentivizing participation by assigning extra credit and devoting class time to completing evaluations are both effective. Knowledge that their responses would make a difference in how a course was taught in the future is also a motivating factor, which is supported by IRPA's phone interview project looking at response rates (Appendix 6).

The committee considered ways to increase the value of the system to students. In addition to recommending that students be given access to all of the numeric results, the committee discussed ways to increase response rates to ensure that the threshold of 70% needed to release results to students is more consistently met. Its recommendations include encouraging instructors to emphasize the value they place in student feedback and set aside class time to complete evaluations. Responding to anecdotal feedback from both students and instructors that an excessive number of items decreases response rates, the committee also considered the length of the instrument. It determined that the number of University-wide items should remain the same, and recommended that the number of College/School/department items be limited to five.

TLTC Initiatives on Evaluating Teaching Effectiveness

Many of the same issues the subcommittee was considering have been considered by the TLTC. The committee learned that the TLTC has been developing a [Mid-Semester Evaluation of College Teaching (MSECT)](), which was piloted in 2017 and 2018 (Donlan, 2019). MSECT is a survey administered through Qualtrics and modeled on the [Fearless Teaching Framework]() (Donlan et al., 2019). It grew out of repeated requests by instructors for a way to evaluate and improve their teaching during the semester. Per the purposes identified for the current course evaluation system, instructors were interested in *formative feedback* that could be used immediately, rather than the next time a course was taught. The tool is intended primarily to help instructors improve their teaching; while instructors could incorporate results into a teaching portfolio, results would not be visible to administrators as with current CourseEvalUM results. Preliminary analysis from pilot data provides evidence that the measure is a valid and reliable assessment of teaching effectiveness. The committee was impressed by the initial results of the pilot, and its recommendations encourage further development and adoption of MSECT and other mechanisms to gather mid-semester feedback on teaching.

Firewall Between Student and Administrator Items

When the University developed its current course evaluation system, the results were treated as "evaluations" of instructors, both rhetorically and in decisions over who should have access to the results. As described above, this approach resulted in two sets of items visible to different audiences with a firewall between them. In the course of its work, and after consulting extensively with the Office of General Counsel, the committee determined that this division is no longer necessary, as results are not "evaluations" used to assess instructor performance.

Student responses are opinions regarding their experience in a particular course. Students are not subject matter experts who can speak to the organization of a course's content, nor are they trained in pedagogy and able to accurately assess an instructor's teaching. Further, there are no standards, rubrics, or specific criteria for students to apply when completing evaluations. Given this, students are not equipped to directly "evaluate" an instructor's performance. Their perspectives can be used by administrators to gain insights into instructors' teaching effectiveness, though in such cases it is the administrator conducting the evaluation, the results/conclusions of which are the actual evaluation to be considered part of the personnel record.

Additionally, all information used in personnel evaluations need not be confidential. There are multiple other sources of information that are not confidential, including numbers and names of publications, syllabi, the number and value of grants, etc. The committee decided that allowing students, instructors, and administrators identical access to numeric results would not invalidate their use in certain personnel processes, and would increase the amount of information available to both administrators and students. It determined, however, that responses to open-ended items should remain confidential and visible only to instructors and administrators as they could contain personally identifiable information and are unaggregated, unit-level data.

Given that results are not performance evaluations, the committee determined that the "course evaluation system" should be renamed to better communicate that it gathers students' perceptions and experiences about a course or instructor, a distinction that current terminology may blur.

Limitations of the Current Course Evaluation System

Based on the reviews of relevant literature and consultation with campus experts addressed above, the committee identified significant concerns with the CourseEvalUM items and their ostensible purpose. The items invite students to speak to themes that they are not in a position to credibly

address. For example, one item asks students to indicate whether "the standards the instructor set for the course were Too Low/Appropriate/Too High," which is better assessed by other experts in the discipline. In addition, many items are not sufficiently specific to yield information that is actionable by an instructor. This lack of specificity is also concerning as vague or imprecise items are more open to bias. The items are also unable to adequately address all four of the system's original purposes:

- **Formative Evaluation**: While results can help instructors improve their teaching, they cannot be used to inform mid-semester adjustments. As addressed above, instructors now have access to more dynamic and timely mechanisms to gather feedback throughout the semester. Further, many of the current items address vague or subjective criteria and do not ask about specific classroom practices associated with effective teaching, making it difficult for instructors to directly address possible deficiencies.

- **Summative Evaluation**: The results remain a potentially valuable measure of teaching effectiveness for use in the APT, AEP, and post-tenure review processes. Yet results are open to misuse, given they yield potentially biased information and tend to measure a single factor (general satisfaction with an instructor). In light of these shortcomings, their use by some as the primary or even sole measure of teaching effectiveness is particularly problematic.

- **Informative Evaluation**: The results remain a valuable resource for students making course selections, though increased completion rates increase information available to students.

- **Outcome Evaluation**: The current system is ill-suited to measuring student learning, which is better addressed by learning outcomes assessments and other mechanisms. Studies have also shown that student learning is not highly correlated with students' perceptions of learning, which are often informed more by other factors (how much the student enjoys the topic, whether the course was required, etc.).

The committee debated at length the purpose of the instrument, eventually determining that it should be redesigned to focus primarily on summative and informative feedback. Surveys should be summative to the instructor and to administrators and serve as one measure of teaching effectiveness to use in evaluating and improving teaching practices. Surveys should also be informative to students, in order to assist them in selecting courses and instructors. The committee determined that the current items are not able to adequately or efficiently meet these goals, and decided to recommend that they be replaced.

FRAMEWORK FOR REPLACEMENT SURVEY ITEMS

The committee decided that the number of items should remain the same as in the current instrument. It considered new items that fall into three conceptual categories:

1. Those designed to provide summative feedback for use in evaluating and improving teaching;
2. Those designed to inform student course decisions; and
3. Those intended to assess teaching assistants.

The committee determined that items in the first category should either assess baseline teaching practices that should be met or identify the utilization of best practices of teaching effectiveness. In discussing the relative balance between these two purposes, the committee considered focusing survey items solely on baseline or core teaching practices that should generally be expected of

every instructor, where consistently low scores can serve as a red flag and inspire discussions between instructors and administrators. It decided that such an approach would only exacerbate concerns that course evaluations are designed to emphasize deficiencies in teaching. It might also suggest that the University's standard for instruction is mere adequacy. Therefore, the committee determined that items addressing baseline practices should be accompanied in approximately equal numbers by items addressing best practices, which presents an opportunity to both identify practices that the University values and encourage adoption of those practices.

In light of the scholarly research discussed above, the committee determined that the work of developing and testing sound replacement items is sufficiently complex that it should be entrusted to those well versed in the scholarship. The committee decided to focus its efforts on identifying constructs that address specific teaching practices and recommend that subject-matter experts be tasked with developing the specific wording associated with each item based on those constructs. Examples of possible wording of survey items associated with most of the constructs can be found in Appendix 8. With the exception of the items intended to inform student course selection, the constructs the committee included in its recommendations are all supported by a large body of literature supporting their connection to learning. The student course selection constructs are based on requests made by students. Following their development, specific survey items would be tested and piloted by the Course Evaluation Advisory Group, IRPA, and subject-matter experts. The final survey items would then be shared with the Senate for its feedback before being implemented. IRPA would also ensure that items are presented in a logical order on the survey, rather than presenting them in the conceptual categories that informed the development of the constructs.

Stakeholder Feedback
The committee circulated its draft recommendations with a range of stakeholders including the Course Evaluation Advisory Group, the Office of Faculty Affairs, administrators, and students. A few stakeholders noted that replacing the current items would disrupt the ability to measure teaching improvement over time. Faculty going up for tenure in the next several years, for example, would have to modify their promotion materials to account for the sudden shift. After consulting with the Office of Faculty Affairs, the committee determined that the APT process could accommodate the change. The AEP process relies more on CourseEvalUM results in some cases, which led the committee to recommend that the University explore ways to provide more holistic reviews of instructional faculty. Adding past results from CourseEvalUM and data from the new survey items to the data warehouse would also facilitate more sophisticated analyses that could smooth the transition, which led to another of the committee's recommendations.

Some stakeholders raised concerns about replacing the current items and about the availability of grade distributions. The committee considered potentially retaining some of the current items for several years to bridge the transition. Stakeholder feedback identified two items in particular as valuable: "I learned a lot from this course" and "Overall, this instructor was an effective teacher." The committee determined not to retain these two items. As noted above, student learning is not highly correlated with students' perceptions of learning, and "effective teacher" is an ambiguous concept subject to significant interpretation. However, several other current items closely align with the proposed constructs, which will facilitate some comparisons between new survey data and CourseEvalUM results. Ultimately, the committee decided that a clean break with the majority of the current items was in the best interests of instructors and administrators. The committee also considered the importance of making grade distributions available to students. Students consistently request them and IRPA annually receives and complies with FOIA requests for grade distributions from third-party websites. Given that students are able to access the information regardless, the committee decided to recommend that results once again be provided directly to students.

RECOMMENDATIONS

I. Guiding Principles & Parameters

1. The University should rename the current "CourseEvalUM" system to better communicate that it gathers students' perceptions and experiences about a course or instructor and does not serve as an evaluation.

2. The University should replace the existing CourseEvalUM items with new survey items that follow these principles:

   a. The number of University-level items should be approximately the same as in the current survey instrument.

   b. The survey instrument should focus on baseline teaching practices, the utilization of best practices of teaching effectiveness, and on a student's perception of teaching effectiveness.

   c. The survey should include only those items that students can reliably answer and should focus on those items where students are the best or only source of data.

   d. The majority of survey items should be designed so that responses can inform actionable changes by the instructor.

   e. The survey items should be written using language that makes clear what is being asked of students.

   f. Students should understand who will be reviewing their responses, in order to inform their thinking as they are filling out the survey.

   g. The survey items should focus on asking students to speak to their own student experience, rather than asking for general feedback or input based on other students' experiences, unless there is a compelling rationale to do otherwise.

   h. The survey items should be relevant for in-person, blended, and online courses.

   i. The survey items should clearly indicate whether they relate to the instructor or the course.

   j. The survey items should be positively worded so that a high score on an item is positive and a low score indicates that adjustments in practices may be needed.

3. The survey item development process should involve a pilot or other mechanism for testing and refining the new items.

4. The Course Evaluation Advisory Group and Office of Institutional Research, Planning, & Assessment (IRPA) should provide an informational report to the Senate on new survey items to gather feedback before implementation.

5. The University should consider ways to ensure that survey results are not utilized as the sole basis for giving teaching awards or for assessing progress towards accreditation standards.

II. Recommendations on Constructs & Items

1. The Course Evaluation Advisory Group and IRPA should work with subject-matter experts to develop new survey items that align with the below constructs that assess teaching effectiveness, inform student registration decisions, provide opportunities for open feedback, and allow for feedback on teaching assistants.

   Constructs that Assess Baseline and Best Practices in Teaching Effectiveness

   - Timely feedback
   - Clear assignment expectations
   - Clear grading expectations
   - Focus on course content in class sessions
   - Value of required texts
   - Climate

   - Instructor support
   - Quality feedback
   - Scaffolding
   - Cognitive engagement and/or rigor
   - Alignment of instruction to assessment

   Constructs that Inform Student Registration Decisions

   - Course satisfaction
   - Instructor satisfaction
   - Time invested

   - Major/non-major
   - Student perception of how much they learned in the course

   Constructs for Open-Ended Feedback

   - Positive Aspects

   - Areas for Improvement

   Constructs Related to Teaching Assistants

   - Climate
   - Timely feedback
   - Effective use of class time

   - Open-ended item on positive aspects
   - Open-ended item on areas for improvement

2. The Course Evaluation Advisory Group, in consultation with the Teaching and Learning Transformation Center (TLTC), should develop a bank of additional items—based on baseline and best practices of teaching effectiveness and literature in the field—that Colleges/Schools and units may include in addition to the University-level items.

3. The University should limit Colleges/Schools and units to a maximum of five additional survey items, which should be developed in consultation with the TLTC.

4. The Course Evaluation Advisory Group should carefully consider the order in which items are presented to students on the survey and whether they should correspond to the order of responses provided in reports available to administrators and instructors.

5. The Course Evaluation Advisory Group should ensure that survey items are clearly identified as applying to either the instructor or to the course.

III. Recommendations on Implementation and Usage of Survey Results

1. The University should encourage instructors to gather mid-semester feedback on their teaching, using tools such as Qualtrics and resources provided by the TLTC.

2. The University should encourage instructors to set aside time in class for students to complete surveys and to explain to students the value and impact of survey responses on teaching practices.

3. The University should make numeric data from survey results available to instructors, administrators, and students. Responses to open-ended items should remain accessible to instructors and administrators only, not students.

4. The University should consider ways to incorporate survey results in information available to students during the course selection process.

5. The Course Evaluation Advisory Group should prioritize efforts to add existing CourseEvalUM data and future survey results to the data warehouse.

6. IRPA should discontinue the practice of including department-wide and College-wide averages across all courses of a given level in survey results.

7. The University should again make course grade distributions available to students.

8. The University should not release survey results from courses with fewer than 5 students and should continue the practice of not releasing results to students if the response rate for a given course is less than 70%.

9. The University should consider how best to ensure that survey results are not utilized as the sole basis for personnel determinations of PTK faculty.

10. The Provost's Office should develop guidance on best practices for utilizing statistical analysis of data from survey results in the Appointment, Promotion, and Tenure (APT) and Appointment, Evaluation, and Promotion (AEP) processes.

APPENDICES

Appendix 1 — Past Senate Action on Course Evaluations
Appendix 2 — Current CourseEvalUM Items
Appendix 3 — Number of Evaluation Items by Unit
Appendix 4 — Big 10 and Peer Institution Research on Course Evaluations
Appendix 5 — Memo from Dr. Sandra Loughlin (November 1, 2017)
Appendix 6 — IRPA Studies on UMD's Course Evaluation System
Appendix 7 — CAWG Snapshot of Student Experiences
Appendix 8 — Sample Item Wording for New Constructs
Appendix 9 — Charge from the Senate Executive Committee

REFERENCES

Doerer, Kristen. 2019. "Colleges Are Getting Smarter About Student Evaluations. Here's How." *The Chronicle of Higher Education*. Retrieved from (https://www.chronicle.com/article/Colleges-Are-Getting-Smarter/245457).

Donlan, A. E., & Byrne, V. L. (2019). "Gathering formative feedback on college teaching: Evidence-based mid-semester evaluation in higher education classrooms." Poster presented at the biennial meeting of the Society for Research on Child Development, Baltimore, MD.

Donlan, A. E., Loughlin, S. M., & Byrne, V. L. (2019). "The Fearless Teaching Framework: A model to synthesize foundational education research for university instructors." *To Improve the Academy, 38*(1), 33-49.

Felton, J., Koper, P. T., Mitchell, J., & Stinson, M. 2008. "Attractiveness, easiness and other issues: Student evaluations of professors on Ratemyprofessors.com." *Assessment & Evaluation in Higher Education*, 33(1), 45-61.

Flaherty, Colleen. 2018. "Teaching Eval Shake-Up." Inside Higher Ed. Retrieved from (https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story).

Lindahl., M. W., Unger, M. L. (2010). "Cruelty in Student Teaching Evaluations." *College Teaching, 58*, 71-76.

Linse, A. R. (2017). "Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees." *Studies in Educational Evaluation, 54*, 94-106.

Supiano, Beckie. 2018. "A University Overhauled Its Course Evaluation to Get Better Feedback. Here's What Changed." *The Chronicle of Higher Education*. Retrieved from (https://www.chronicle.com/article/A-University-Overhauled-Its/243803).

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, 54*, 22-42.

Summary of Past Senate Action on the Topic of Course Evaluations:

In July 2002, The Educational Affairs Committee was charged with reviewing a proposal from Lilly-CTE Fellows to establish a University policy on the evaluation of teaching (Senate Doc #01-02-63). Senate Chair Kent Cartwright sent a memo to John Pease, Chair of the Educational Affairs Committee, asking the committee whether it would like to examine the proposal in depth or forward it to a joint task force of the Senate & Academic Affairs for further study. The memo detailed specific issues and questions that should be considered, and the proposal from Lilly-CTE was attached.

In November 2002, the Educational Affairs Committee responded to the SEC, stating that it had decided not to make a formal recommendation regarding the Lilly-CTE proposal for the Establishment of a University Policy on the Evaluation of Teaching. It suggested that a Task Force be created to look into this issue further.

On January 14, 2003, the SEC reviewed the memo from the Educational Affairs Committee and voted to develop a proposal for a Task Force.

The Joint Task Force on Course Evaluations and Teaching was appointed by the Office of the Provost and the University Senate. The Task Force was charged during in the spring of 2003.

The Task Force met during the summer and fall of 2003. It presented an interim report in February 2004. One of the recommendations from this report became a resolution for a university-wide requirement for student evaluations in all undergraduate and graduate courses.

The University Senate passed the resolution on May 3, 2004, mandating a university-wide requirement for student evaluations in all undergraduate and graduate courses. Senate Doc 02-03-39 stated "we recommend that there be a university-wide requirement for student evaluations in all undergraduate and graduate courses."

Following the passage of the resolution, the SEC updated the original charge to the Task Force in September 2004. The Task Force sent a draft response to the updated charge and a draft of their final report to the SEC for its meeting on January 19, 2005 (draft report dated January 12, 2005). The draft report detailed a set of six recommendations calling for, in part, a university-wide course evaluation system (web-based), a set of universal evaluation questions, and that a portion of the evaluation results be made public to the students. On January 19, 2005, the SEC met to review the response from the Task Force to the updated charge and draft report.

The Task Force compiled its Final Report in April 2005. This report contained seven recommendations on how the academic community could enhance its capabilities to assess and improve curriculum and instruction. The Task Force members unanimously agreed that a university-wide course evaluation requirement and system should be adopted.

The SEC met on September 13, 2005, and approved a consultation between Senate Chair Berlin and the Task Force to draw certain recommendations from the final report to be presented as actionable items to the Senate, along with a report from Provost Destler on implementation.

The SEC met on November 1st and voted to invite the Chair of the Task Force to the next meeting, along with the lawyer who had been advising them.

The Task Force presented its report and recommendations to the SEC on November 15, 2005. The SEC decided that Chair Berlin would work with the Task Force to revise the language of its recommendations.

The Task Force presented a revised document to the SEC on November 29, 2005. The SEC voted to approve the Task Force's document for the December Senate agenda.

On December 12, 2005, the Chair of the Task Force, Dennis Kivlinghan, presented the actionable recommendations (Recommendations for the Implementation of Web-based Student Course Evaluations, Senate Doc #02-03-39). He explained that the nine recommendations were principles for implementing web-based course evaluations. The recommendations would be implemented through the Provost's Office.

Chair Berlin sent a memo to President Mote on December 15, 2005, stating that the Senate had approved the Recommendations for the Implementation of Web-based Student Course Evaluations.

President Mote accepted the recommendations on December 21, 2005. He stated that there remain significant issues for full implementation, both in timing and in framing the questions, and gave suggestions for how to move forward.

Chair Berlin reported to the SEC about Dr. Mote's letter at the SEC meeting on January 24, 2006. Berlin noted that the Provost had formed an implementation committee. VP and CIO Jeff Huskamp presented an informational summary of technology issues relating to the implementation of web-based student evaluations to the SEC on February 28, 2006.

Sharon La Voy Chaired the Provost's Student Course Evaluation Implementation Committee and she presented the committee's university-wide questions for online student evaluations at an SEC meeting on March 14, 2006. The questions had been reviewed by the Council of Deans. The SEC made changes, and La Voy presented a final set of questions on April 11, 2006. The SEC voted to place the questions on the April 24th Senate agenda as an informational item.

The Provost and the Implementation Committee presented the questions for the web-based evaluation instrument. The Provost explained that the Senate would not be asked to approve the questions but to provide feedback. He confirmed that responses to the set of questions for APT would not be made public. The Provost emphasized that he would require a 75% participation rate before results for a course would be published. He explained that the new system would be fully implemented in the fall of 2007.

On May 29, 2007, Chair Montgomery sent a memo to VP and CIO Jeff Huskamp expressing disappointment that implementation of the online course evaluations had been halted due to a technical problem of putting a questionnaire that includes universal, college, departmental, and professorial questions online. The SEC passed a resolution stating that a questionnaire with only universal questions be available online campus-wide in the Fall 2007 semester (Senate Doc. 06-07-56).

On August 31, 2010, the SEC charged the APAS Committee with review of Re-evaluation of the Student Teach Evaluations at UMD (Senate Doc. 10-11-36) proposed by Denny Gulick (Mathematics Professor and Past Chair, 1998-1999, of the Senate).

The APAS Committee reviewed this charge in Fall 2010. At its September 3, 2010, meeting the committee reviewed background history on this topic as provided by the Senate Office. It also researched peer institution procedures for course evaluations, off-campus course evaluation services, and potential legal concerns. During the course of its review, the APAS Committee read articles on the subject of teacher evaluations and consulted with members of the Office of Institutional Research Planning & Assessment (IRPA). Following deliberation, the APAS Committee voted, at its December 17, 2010, meeting, in favor of recommending that the CourseEvalUM system continue to undergo development

with the guidance of a governing body that is formulated in a manner consistent with the principles of shared governance. The APAS Committee's report also outlined a number of specific subjects that warranted further attention, including the recommendation that more detailed consideration should be given to how CourseEvalUM could be modified to better satisfy student needs. Additionally, the APAS Committee strongly endorsed the urgency for the addition of unit-specific questions, including course-specific and instructor-specified questions to the CourseEvalUM system.

On January 28, 2011, the SEC reviewed the APAS Committee's report and voted to forward the report to the Senate as an informational item. The SEC also voted to send a letter to the Provost requesting administrative action and a report describing actions taken by September 1, 2011. The report was presented as an informational item at the February 9, 2011 Senate meeting.

On September 1, 2011, Provost Ann Wylie sent a response to the SEC regarding the APAS Committee's report. This letter discussed the report and offered a number of recommendations (See attached letter).

On October 13, 2011, the APAS Committee wrote a response to the Provost's letter requesting more information on the implementation of unit-specific questions. The SEC forwarded this letter to the Provost on October 28, 2011.

On January 18, 2012, the SEC received a response from the Provost regarding the October 28, 2011 memo. The response included information on how the priorities for developing the CourseEvalUM system were decided and the consideration given to instructor-specified and course-specific questions. The APAS Committee reviewed this letter on February 27, 2012.

Prepared by the Senate Office – February 2017

Current Course Evaluation Items

Utilizing a universal set of course evaluation questions allows both students and academic administrators to make more meaningful and consistent comparisons among courses and their instructors. Evaluation items fall into groups based on who has access to the results as explained below. Unless otherwise noted, items are answered on the following scale:

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Not Applicable

Administrator Items

Results from these items are seen by instructors and authorized campus administrators. Administrators use evaluation results to help them assess and improve faculty performance. Because faculty performance review is a personnel function, the evaluation feedback used in those decisions is confidential by Maryland law. For this reason, students and administrators see responses to different sets of items from the evaluation.

1. The instructor treated students with respect.
2. The instructor was well-prepared for class.
3. The course was intellectually challenging.
4. The standards the instructor set for students were... (Too Low, Appropriate, Too High)
5. I learned a lot from this course.
6. Overall, this instructor was an effective teacher.
7. How does this course fit into your academic plan or course of study? (CORE or General Education Requirement, Major/Certificate/Minor/Program Requirement, Elective)
8. Additional comments, e.g. about course content/materials, teaching style, etc. (free-text item)

Student Items

Results from these items are seen by instructors and students. Students use evaluation results to aid them in choosing courses each term.

1. The instructor was effective in communicating the content of the course.
2. Course guidelines were clearly described in the syllabus.
3. The required texts (e.g., books, course packs, online resources) helped me learn course material. (added Fall 2017)
4. The instructor was responsive to student concerns.
5. The instructor helped create an atmosphere that kept me engaged in course content.
6. Based on the quality of my work in this course, the grades I earned were… (Too Low, Appropriate, Too High)
7. Given the course level and number of credits, the workload was… (Too Low, Appropriate, Too High)
8. How much effort did you put into the course? (Little, Moderate, Considerable)

Teaching Assistant Items

Results from these items are seen by teaching assistants, instructors teaching with the TA, and campus administrators.

1. The teaching assistant (TA) treated students with respect.
2. The teaching assistant (TA) was well-prepared for class.
3. Overall, this teaching assistant (TA) was an effective teacher.
4. Additional comments, e.g. about the discussion/lab/studio section, TA's teaching style, etc. (free-text item)

CourseEvalUM - Number of Evaluation Items

Unit	Course Questions		Instructor Questions		Total without TA	TA Questions		Total with TA
	Single Selection	Open Ended	Single Selection	Open Ended		Single Selection	Open Ended	
University-wide	*8*	*1*	*7*	*0*	*16*	*3*	*1*	*20*
ARHU	0	0	0	0	16	5	0	25
ARHU-English	3	2	1	0	22	1	0	32
BMGT	1	0	0	0	17	0	0	21
BSOS	0	0	10	0	26	0	0	30
BSOS-Psychology	6	0	0	0	32	0	0	36
EDUC	1	2	4	0	23	0	0	27
ENGR	16	0	0	0	32	0	0	36
INFO	3	0	2	0	21	0	0	25
JOUR	2	2	2	3	25	0	0	29
PLCY	12	0	10	0	38	0	0	42
UGST-College Park Scholars	3	4	0	0	23	0	0	27

Updated 8/5/2017

Initial Peer Research on Course Evaluations
(Compiled by Ben Bederson and TLTC)

Commonalities among the Big 10
- Each university has a set of Standard/Cross-campus questions which are mandatory. Then there are departmental questions, and then the faculty can select questions.
- Some questions are multiple choice and some are open-ended. Scales differ.
- As universities move course evaluation online they have been developing strategies for incentivizing online completion.
- All campuses allow faculty to access their own student course feedback
- Reporting methods:
    - OSU provided class and cumulative reports
    - Illinois also reports longitudinal data by faculty member and/or course.
- Wisconsin was the only school identified whose course evaluations were linked to state/system-wide evaluations of teaching and hiring/promotion/salary-raises decisions.

Indiana University Bloomington
- 2014 Online Course Evaluation Template:
    - They experimented with passive vs. active voice
    - When piloting, found that students receiving F's were least likely to respond, and students receiving A's were most likely to respond.
    - Add a "not applicable" option
    - Graduate students have higher instructor ratings that undergraduates
    - Depending on the question, first year students answer differently than older students.
    - Professional graduate students find the some questions to not fit their in-class experience.
    - "student course evaluations should be judged in relation to contextual characteristics, such as class size, level, major requirement status, and other factors that systematically influence student perceptions."
- Other Information:
    - https://academics.iusb.edu/institutional-research/online-course-questionnaire.html

Michigan State
- Evaluation summaries are available to students: https://sirsonline.msu.edu/FAQ.asp
    - "Student Opinion of Courses and Teaching (SOCT) collects feedback from undergraduate students enrolled in classes taught by MSU faculty during fall and spring semesters. SOCT surveys are not collected for summer courses or any courses taught by graduate assistants. SOCT questions were developed to gather information that may be helpful to students when selecting courses and faculty members in those courses. The aggregate results of this survey are updated at least twice a year and are available to the MSU community."

- Encourage faculty to do mid-course evaluations
- Tools for evaluating online courses
- Example of how one college uses eval results in promotion and hiring https://natsci.msu.edu/faculty-staff/policies-procedures/evaluation-policy-and-resources/teaching-evaluation-guidelines/

Northwestern University
- Not available online. Will need to request information from the below office:
- http://www.northwestern.edu/ses/faculty-instructors/ctecs/running-instructor-ctec-reports.html

Ohio State University
- Standard Form:
  - https://registrar.osu.edu/sei/seiitems.pdf
  - Not flexible for team teachers
- Optional Feedback on Your Instruction (FYI) program: http://ucat.osu.edu/professional-development/fyi/
  - Only for instructor use
  - Flexible for team teachers
- Reporting: Instructors can get a report for just their one class or a report of "Overall Scores" across all courses the professor has taught
- Has student view

Pennsylvania State University
- All mandatory & approved questions: https://www.srte.psu.edu/SRTE_Items/
  - University required:
    - A1. Are you taking this course as an elective? (If uncertain, omit.)
    - A2. What grade do you expect to earn in this course?
    - A3. Rate the overall quality of this course.
    - A4. Rate the overall quality of the instructor.
  - Then, organized by Departmental questions, Instructor-selected questions, University open-ended questions, etc.
- Student Rating Teaching Effectiveness: https://www.srte.psu.edu/
- NOT available to students. "SRTE results are considered part of faculty members' personnel records so access is restricted to the faculty member and administrators."
- Faculty beliefs about encouraging student participation: http://www.schreyerinstitute.psu.edu/IncreaseSRTERespRate/

Purdue University
- Senate Teaching Evaluation Conceptual Overview
  - University required:
    - Overall, I would rate this course as: Excellent - Good - Fair - Poor - Very Poor.
    - Overall, I would rate this instructor as: Excellent - Good - Fair - Poor - Very Poor.

- "All course evaluations include 8 standard questions, the two University "Core" items, four demographic questions used for research purposes, and two written prompts for student feedback."
- https://www.purdue.edu/cie/IDP/courseevaluations.html

Rutgers University–New Brunswick
- Increase response rate by using a midterm informal feedback form: https://ctaar.rutgers.edu/sirs/participation.html
- Administer both paper and online surveys
- How to interpret responses: https://ctaar.rutgers.edu/sirs/guidelines.html
- Online sample: https://ctaar.rutgers.edu/sirs/osirsPreview.html
    - 1. The instructor was prepared for class and presented the material in an organized manner. N/A, Strongly disagree---Strongly agree
    2. The instructor responded effectively to student comments and questions. N/A, Strongly disagree---Strongly agree
    3. The instructor generated interest in the course material. N/A, Strongly disagree---Strongly agree
    4. The instructor had a positive attitude toward assisting all students in understanding course material. N/A, Strongly disagree---Strongly agree
    5. The instructor assigned grades fairly. N/A, Strongly disagree---Strongly agree
    6. The instructional methods encouraged student learning. N/A, Strongly disagree---Strongly agree
    7. I learned a great deal in this course. N/A, Strongly disagree---Strongly agree
    8. I had a strong prior interest in the subject matter and wanted to take this course. N/A, Strongly disagree---Strongly agree
    9. I rate the teaching effectiveness of the instructor as: N/A, Poor----Excellent
    10. I rate the overall quality of the course as: N/A, Poor----Excellent
- Paper sample: https://ctaar.rutgers.edu/images/SIRS_form.jpg
- How to add additional questions: https://ctaar.rutgers.edu/sirs/addQuestions.html
- Information hub: https://www.purdue.edu/cie/IDP/courseevaluations.html
- Use in faculty portfolio: http://senate.rutgers.edu/bestprac.html

University of Illinois at Urbana–Champaign
- Full catalogue of all question items: https://citl.illinois.edu/docs/default-source/default-document-library/icescatalog.pdf?sfvrsn=0
- Info hub: https://citl.illinois.edu/citl-101/measurement-evaluation
- Reporting: Each semester or longitudinally by course or instructor: https://citl.illinois.edu/docs/default-source/ices-documents/sample-longitudinal-profile.pdf?sfvrsn=2
- Paper example: Front, Back
- Mid-semester feedback surveys are encouraged: https://citl.illinois.edu/citl-101/measurement-evaluation/teaching-evaluation/ief
- Online:

- ICES Online allows 23 rated items and 6 open-ended items in addition to 3 global items. Faculty can write their own open-ended items. Faculty are not allowed to write any rated items, but we are continually expanding the item pool and welcome suggestions for new items.
- Opportunity to tailor for team teaching
- Have the option to not release to the department.

University of Iowa
- How to administer to prevent bias and increase responses: https://teach.its.uiowa.edu/ace-online-best-practices
- Global Items:
    - 101.This course is well planned and organized.
    - 102. The content of this course is valuable.
    - 103. This is a worthwhile course.
    - 104. Overall, this is an excellent course.
    - 105. I learned more in this course than in most other college courses I have taken.
    - 106. I learned a great deal in this class.
    - 107. I am motivated to do my best work in this course.
    - 108. This instructor is effective in teaching the subject matter of this course.
    - 109. Overall, this instructor is an effective teacher.
    - 110. This instructor is an excellent teacher.
    - 111. I would recommend a course taught by this instructor to other students.
- Item pool: https://teach.its.uiowa.edu/sites/teach.its.uiowa.edu/files/ace_item_pool.pdf
- Information hub: https://teach.its.uiowa.edu/technology-tools/ace-online-course-evaluations

University of Minnesota
- Mostly using paper surveys.
- Information Hub: https://oms.umn.edu/srt/
- Sample paper form: https://oms.umn.edu/departments/srt/answerSheets.php
- HOW responses are used: http://policy.umn.edu/education/teachingevaluation
    - "When used for salary, promotion, and tenure decisions, information from student ratings should be used in conjunction with other relevant metrics to assess instructional effectiveness."
- FAQ: https://oms.umn.edu/departments/srt/about.php
  The SRT Course Items are:
    - I have a deeper understanding of the subject matter as a result of this course.
    - My interest in the subject matter was stimulated by this course.
    - Instructional technology employed in this course was effective.
    - The grading standards for this course were clear.
    - I would recommend this course to other students.

- Approximately how many hours per week did you spend working on homework, readings, and projects for this course?
  - 0-2 hours per week
  - 3-5 hours per week
  - 6-9 hours per week
  - 10-14 hours per week
  - 15 or more hours per week

University of Nebraska–Lincoln
- https://canvas.unl.edu/courses/1/quizzes/7?module_item_id=60
- Little available information online
- Faculty can add additional questions: http://cehs.unl.edu/cyaf/course-evaluations-0/
- Housed in Blackboard
- Components: Command and Connection:
  - http://www.unl.edu/gradstudies/current/news/using-student-evaluations

University of Michigan
- "When core templates need creating or modifications, instructors working with their evaluation coordinators can design their core evaluation templates by selecting the **questions from our Question Catalog** A maximum of 30 rating questions and 5 open-ended comment questions is the limit.
- Core questions:
  - Text, Level, Dimension
  - I had a strong desire to take this course., Course, Student Motivation
  - As compared with other courses of equal credit, the workload for this course was (SA=Much Lighter, A=Lighter, N=Typical, D=Heavier, SD=Much Heavier)., Course, Perceived Workload
  - This course advanced my understanding of the subject matter., Course, Self-assessed Learning
  - My interest in the subject has increased because of this course., Course, Impact on Students
  - I knew what was expected of me in this course. (SA=Almost Always, A=Frequently, N=Sometimes, S=Occasionally, SD=Hardly Ever)., Course, Course Organization
  - The instructor seemed well prepared for class meetings. (SA=Almost Always, A=Frequently, N=Sometimes, S=Occasionally, SD=Hardly Ever), Instructor, Course Organization
  - The instructor explained material clearly. (SA=Almost Always, A=Frequently, N=Sometimes, S=Occasionally, SD=Hardly Ever), Instructor, Instructor Clarity
  - The instructor treated students with respect., Instructor, Classroom Climate
- Midterm and Final course evals
- http://ro.umich.edu/evals/#FS_Templates_Questions

University of Wisconsin–Madison
- First to address Climate as well as "information on key initiatives not typically captured by evaluations, such as alignment with the campus Essential Learning Outcomes"
- Faculty and departments can add questions
    - Moved departments in fall 2016: "Testing and Evaluation (T&E) no longer offers online course evaluations."
      https://testing.wisc.edu/onlinecourseevals.html
    - Now held under Teaching and Learning:
      https://learnuw.wisc.edu/toolbox/aefis.html
        - https://provost.wisc.edu/assessment/digital-course-evaluation-surveys.htm
- INFO HUB: https://testing.wisc.edu/standardizedcourseevals.html
- Wisconsin was the only school identified whose course evaluations were linked to state/system-wide evaluations of teaching and hiring/promotion/salary-raises decisions.
- Critique from student newspaper:
  https://badgerherald.com/news/2015/04/30/course-evaluations-get-a-failing-grade-in-terms-of-effectiveness/

Other Institutions (Non-Big 10, Peer)

Ball State University
Contact: James A. Jones, PhD
Director, Research and Academic Effectiveness
Ofc of the Assoc Provost & Dean, Univ College
"Generally, our response rates have been around 50% or better for the campus overall. There is a lot of variability among classes, instructors, departments, and colleges, however. The class ratings for the items tend to average around 4.2 on a 5-point scale with 1 being the negative end and 5 the positive one. This indicates the fear that instructors raised that only the students with negative opinions would complete the ratings does not appear to be a problem. There is also very little relationship between response rate and ratings received, indicating that having lower response rates than what we had compared to paper administration is not causing obvious harm.

I have attached what we use for core questions. Each department has the option of having additional questions included, and about half our departments do so. There is also a period of time about a week before the evaluation period opens that we allow instructors to add their own questions."

George Mason University
Contact: Melissa A. Broeckelman-Post, PhD, Assistant Professor and Basic Course Director, Department of Communication
"The committee started our process by reviewing other forms and doing an extensive literature review to try to identify whether there are existing high-quality (reliable and valid) measures of teaching, and they were unable to find any measures for which there was an evidence of validity. So, they decided to take on a several year process in which we first used literature to try to identify categories of criteria related to effective teaching with several examples of the types of items that might be included, and we're now surveying deans, directors, and faculty evaluation committees about which categories they think are important as well as how they use the student evaluation of teaching forms. This is about the point where I joined the committee. Our next steps are to survey faculty and students about which categories they think are important, after which we'll develop some actual potential survey items and do another faculty and student survey.

A year or two ago, though, one of my colleagues who teaches advanced quantitative methods here was talking about this issue and said that the evaluation process developed at K-State is (in her opinion) one of the most reliable and valid instructor evaluation forms in use, in part because it also accounts for students' own engagement in the course. I haven't dug in much further than that (and our committee didn't find this during their earlier search), but I plan to soon-- for now, here's the link that my colleague shared at that time: http://www.k-state.edu/tlc/course-evaluation/forms.html. It's probably worth a conversation with their Teaching & Learning Center to learn more."

Illinois State University
Contact: Cheri J. Simonds, Professor, Co-Director of Communication as Critical Inquiry School of Communication, Illinois State University
"Here is our departmental course evaluation. I was on the team that created this instrument and it is based in instructional communication theory and research. We have found that students are much more thoughtful and constructive in their feedback for instructors. I hope you find this useful."

Kansas State University
http://www.k-state.edu/tlc/course-evaluation/forms.html
"Our signature service, the IDEA **Student Ratings of Instruction** instrument (SRI) is like no other system available for translating course feedback into actionable steps to improve learning. The SRI system is supported by extensive research, controls for extraneous circumstances (e.g. class size, student motivation), and provides comparative scores. Faculty and administrators can easily integrate data into program planning, decision making, accreditation and institutional review processes. Through our partnership with Campus Labs, we offer a paperless solution with an intuitive, mobile-friendly interface."

UNIVERSITY OF
MARYLAND

ROBERT H. SMITH
SCHOOL OF BUSINESS

OFFICE OF
TRANSFORMATIONAL
LEARNING

TO: Dr. Phil Evers, Chair of Academic Procedures and Standards Committee
FROM: Dr. Sandra Loughlin, Director, Office of Transformational Learning, Robert H. Smith School of Business
DATE: November 1, 2017
RE: **Current Research on Use of Student Evaluation to Assess Teacher Effectiveness**

Dear Phil,

I am responding to APASs request for a review of the literature on student evaluation of teaching (SET) in higher education. This memo is a brief summary of the extant literature, with an emphasis on studies of the highest methodological caliber. In evaluating these studies and drawing conclusions, I consulted with other experts in learning and educational measurement, including Drs. Patricia Alexander, Gregory Hancock, Joshua Polanin, Elizabeth Richey, and Alice Donlan.

Please note that this analysis assumed that the purpose of SET is to primarily to assess teacher effectiveness, rather than students' satisfaction of a course. If the committee determines that purpose of the student evaluation is satisfaction, rather than an indicator of teacher effectiveness, many of these findings and recommendations are irrelevant.

Findings
- There is a significant literature on SET, however the majority of the studies use poor methods, yielding highly suspect and ungeneralizable findings. This memo only includes studies with rigorous methodologies.
- When used as the only/primary source of data, SET it is very poor indicator of teacher effectiveness. It should only be used in combination with other measures (e.g., peer evaluation of course materials, assignments, and assessments).
- Current SET instruments, including UMDs SET, routinely ask students to assess factors for which they are a poor source of data (e.g., whether the instructor is knowledgeable in his/her area).
- A significant body of research shows that SET is not associated with student learning.

- There are a few rigorous studies examining the degree to which SET is influenced of factors unrelated to teaching effectiveness (e.g., instructor gender or the type of course). These studies consistently show the teaching irrelevant factors have small, but significant influence on SET.
- Although not a focus of my research, I found that there is precedent for instructors to sue universities for wrongful termination based on reliance on SET data (Maffly, 2011). In light of the findings that SET is a poor indicator of teaching effectiveness, this may be a source of concern.

Recommendations
- Empirically test the degree to which UMDs SET measures teacher effectiveness.
- Under advisement of experts in educational measurement, consider revising the current SET instrument, administration, and data usage.
- Investigate the degree to which UMD departments and schools use SET as the only/primary source of data on teacher effectiveness for promotion, retention, tenure, merit pay, and other decisions like teaching awards.
- Provide training to administrators and faculty on the appropriate interpretation and use of SET data.

There is significant research on the value of student evaluations of teaching (SET) in higher education. Indeed, a cursory examination of the literature reveals literally thousands of studies on the subject, which draw widely varying conclusions. The high variance in study outcomes is likely related to the equally wide variance in the quality of the study design; unfortunately, much of the SET research is methodologically poor, suffering from serious threats to validity and generalizability (Hornstein, 2017; Linse, 2017; Stark & Freishtat, 2014; Wieman & Gilbert, 2014). The following is a brief discussion of the extant literature, with an emphasis on recent, methodologically sound studies.

Teacher effectiveness is a latent construct and measurement of it requires multiple sources of data. A meaningful assessment of teaching effectiveness would draw from at least five sources (Berk & Theall, 2006; Spooren, Brockx, & Mortelmans, 2013; Weiman, 2015).

- *Peer evaluation* of course materials, assessments, and assignments to determine if the course is current, rigorous, and in line with program curriculum.
- *Pedagogy expert evaluation* of class sessions and course design to determine if the instructor is using instructional practices that are predictive of student learning.
- *Direct assessment of learning* to determine the degree to which students achieved the learning goals established for the course and succeed in follow-on courses.
- *Instructor-generated portfolio* that documents how the instructor uses student learning data and other sources of feedback to improve instruction and student outcomes.
- *Student evaluation of teaching* to understand students' experience and perception of the instructor.

Although important to capture, student perception alone is an insufficient measure of teaching effectiveness, because students are not positioned to provide valuable information on many aspects of instruction (Langbien, 2008; Linse, 2017; McKeachie, 1997; Stark & Freishtat, 2014). Despite this fact, SET instruments routinely include questions for which students are a poor source of data (e.g., *The instructor is knowledgeable in his/her area*; Becker, Bosshardt, & Watts, 2012; Hornstein, 2017). This inappropriate reliance on student opinion on areas best assessed by others is evidenced in all commonly used SET (for a listing of SET instruments, see Spooren, Brockx, & Mortelmans, 2013).

Existing SET instruments are flawed measures of teacher effectiveness and should not be used as the only/primary source of data.

A significant body of research shows that SET does not explain variance in learning outcomes.

- 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*, *54*, 22-42.
  - Rigorous analysis of data from 97 multisection courses that include student evaluations and course outcome measures. **NOTE:** This study reanalyzed data from previously conducted studies, which exposed significant methodological flaws in prior research. Unfortunately, most reviews of SET have used the prior, poorly-designed meta-analyses and have drawn suspect conclusions regarding the concurrent validity of SET (e.g., [Spooren, Brockx, & Mortelmans, 2013]()).
  - **FINDING:** On average in the 97 studies, SET explained >1% of variance in course grades.
  - **FINDING:** Small sample-size studies tended to show correlations between SET and learning, but large sample-size (and presumably more robust) studies did not.
  - **CONCLUSION:** Data suggests no meaningful correlation between SET and learning outcomes.
- Carrell S. E., & West J. E (2010). [Does professor quality matter? Evidence from random assignment of students to professors](). Journal of Political Economics, *118,* 409–432.
  - Methodologically rigorous, large-scale study (n = 12,568) with random assignment of students to courses. Using an introductory course with different instructors, the authors examine the relation between instructor factors (i.e., rank, years teaching, degree), student evaluations, and student performance in contemporaneous and follow-on courses.
  - **FINDING:** Instructor factors were negatively correlated with performance in contemporaneous course, but positively correlated with follow-on courses.
  - **FINDING:** Evaluations were positively correlated with contemporaneous course, but negatively correlated with follow-on courses.
  - **FINDING:** Contemporaneous and follow-on course performance were negatively correlated.

- **CONCLUSION:** Suggests evaluations predict course performance, but not significant learning as measured by later performance. Suggests that expert instructors (i.e., those with higher rank, years teaching, and degree) do a better job preparing students for success in follow-on courses. Also suggests we need to look beyond the course in question to really measure student learning.
- Weinberg, B. A., Fleisher, B. M., & Hashimoto, M. (2008). Evaluating teaching in higher education. *Journal of Economic Education.*
    - Using a clever design, SET and course grades were collected from 4,111 students in a foundational course and two follow-on courses.
    - **FINDING:** SET and current course grade were consistently correlated.
    - **FINDING:** When controlling for grade in prior course, SET and course grade were no longer associated.
    - **CONCLUSION:** Findings suggest a weak relationship between SET and learning.

The current literature suggests that SET is influenced by factors unrelated to teaching effectiveness, such as course type and instructor gender. **NOTE:** There is a considerable literature on gender bias showing mixed effects, but most studies in this space are poorly designed. The studies included here are experiments and represent the most rigorous level of research available on the subject.
- Uttl B, Smibert D. (2017) Student evaluations of teaching: teaching quantitative courses can be hazardous to one's career. *PeerJ* 5:e3299
    - Using a rigorous and appropriate analytical design, the study examined 35,538 courses to determine whether qualitative and quantitative courses demonstrated different patterns of SET.
    - **FINDING:** On average, quantitative courses were associated with significantly lower SET scores.
    - **FINDING:** Distribution of SET also differs by course type, with SET in quantitative courses approximating the normal distribution and SET in qualitative courses show a negative skew and high mean ratings. If cut scores for SET are arbitrarily set across schools and hold constant, instructors of quantitative courses may be a higher risk of being labeled as unsatisfactory.
    - **CONCLUSION:** SET may have a disproportionate, negative impact on instructors of quantitative courses, which may lead to negative repercussions for tenure, promotion, and/or merit pay.

- MacNell, L., Driscoll, A., & Hunt, A. N. (2015). [What's in a name: exposing gender bias in student ratings of teaching](). *Innovative Higher Education*, *40*(4), 291-303.
  - Clever experimental study in which students were randomly assigned to two online instructors, both of whom operated both male and female identities. This created a 2x2 condition. Instructors coordinated to write similar bios, use the same assignments and grading scale, return grades at the same time, and use the same level and pattern of interpersonal interaction. SET was written to include behaviorally-worded items (e.g., grades were returned promptly).
  - **FINDING:** Students who perceived their instructor to be male gave higher ratings of fairness, promptness, and praise, even though both instructors used the same assignments and grading scale, returned grades at the same time, and used the same level and pattern of interpersonal interaction. There were no significant differences based on instructors' actual genders.
  - **CONCLUSION:** Suggests that SET may be biased in favor of men *even when items are behavioral.*
- Arbuckle, J., & Williams, B. D. (2003). [Students perceptions of expressiveness: Age and gender effects on teacher evaluation](). *Sex Roles,* 49, 507-515.
  - A laboratory experiment in which students (n=352) were shown "slides of an age- and gender-neutral stick figure and listened to a neutral voice presenting a lecture and then evaluated it on teacher evaluation forms that indicated 1 of 4 different age and gender conditions (male, female, 'old,' and 'young')" [11, p. 507]. All students saw the same stick figure and heard the same voice, so differences in SET could be attributed to students' perceptions of the age and gender of the instructor.
  - **FINDING:** When students were told the instructor was young and male, students rated the instructor higher than for the other three combinations, especially on "enthusiasm," "showed interest in subject," and "using a meaningful voice tone."
  - **CONCLUSION:** SET may be biased in favor of males and younger instructors.
- Leventhal, L., Perry, R. P., & Abrami, P. C. (1977). [Effects of lecturer quality and student perception of lecturer's experience on teacher ratings and student achievement](). *Journal of Educational Psychology*, *69*(4), 360.
  - Experimentally manipulated lecturer quality and students' beliefs about instructors' experience. Students watched videos of high- or low-quality lecture on the same content (e,g., in the low-quality lecture, the instructor

stammered, was disorganized, was less enthusiastic) and students were told the instructor had a lot or very little experience. This created a 2x2 condition. Students were then directed to rate instructional quality and took a post-test.

- **FINDING:** Lecturer quality was more strongly associated with ratings than it was with student achievement.
- **FINDING:** Students' beliefs about the instructor's level of experience affected the relations between quality and ratings.
- **CONCLUSION:** Suggests SET is influenced by factors not associated with learning, such as presentation quality and students' perceptions about the instructor's level of experience.

The degree to which UMD's SET measures teacher effectiveness is unknown.
- To date, UMD has not conducted studies linking SET data to factors associated with teacher quality, such as learning outcomes in the current course or follow-on courses, evaluation of pedagogy by a teaching expert, or evaluation of course or materials by a peer.
- To date, UMD has not conducted studies linking SET to factors that may inappropriately skew data, such as the gender, age, and race of the instructor; whether the course is qualitative or quantitative; or whether the course is required or elective. Ideally, a statistical model would be developed to control for these factors.

UMD should consider empirically testing the degree to which the existing SET measures teacher effectiveness. This would entail linking historical SET data to learning outcomes and other measures of teacher effectiveness, where possible (e.g., peer evaluation of materials, expert evaluation of pedagogy and course design); investigating the relationship between SET and potential sources of bias; examining the factor structure in the existing measure to determine if the instrument has differential functionality in colleges/programs; assessing the degree to which the instrument as a whole, and at the item level, explains variance in learning outcomes; and examining the stability of SET scores for instructors over time. In addition, new studies involving SET could be devised, such as identifying the correlation between SET and self-reported student satisfaction or SET and self-reported student motivation.

Alternatively or in addition to studying the existing SET, UMD should consider revising it. While there is no empirical evidence to indicate whether UMDs SET instrument appropriately measures teaching effectiveness, there is evidence that the portion of current instrument that generates data shown to departments and instructors

includes items for which students may not be the best source of data. For instance, UMD students are currently asked to rate the degree to which the course was *intellectually challenging* and whether the instructor *set appropriate standards* for students. These aspects of effective teaching are best assessed by a faculty peer with knowledge of the intellectual rigor and standards necessary for the course in light of follow-on courses and the demands of the field. Students do not, nor should they be expected to, have this knowledge. UMD students are also asked to assess the degree to which they *learned a lot* from the course. Learning is best assessed by a direct measure such as performance on the final exam/project and because students are notoriously poor judges of their own learning (Tai, Klayman, & Hastie, 2008).

In light of previous research and the current configuration of UMDs SET, revision to the current instrument may be warranted. In this effort, UMD should leverage the expertise of learning and educational measurement experts. Creating a valid, reliable measure of a latent construct such as teacher effectiveness is a complex and difficult process that requires considerable training in educational and psychological measurement (Berk & Theall, 2006; Gall, Gall, & Borg, 2003). Given the potentially high-stakes use of SET data, the instrument development process should reflect a rigorous approach to measurement design (for a brief overview of the process, see Korb, 2017).

The creation of the Student Evaluation of Teaching in Medical Lectures SETMED-L (Mueller et al., 2017) is a good example of the correct approach to developing a SET instrument. Of particular note is the fact that the authors grounded the instrument a theoretical framework of effective teaching (i.e., the Stanford Faculty Development Program). While I would suggest grounding a new UMD SET in the Fearless Teaching Framework rather than the Stanford program, starting with a research-based framework is a critical first step toward creating a valid, reliable measurement of teaching or learning.

The only significant methodological concern with the creation of the SETMED-L is the fact that, like all other SETs found in the literature, the instrument includes some items for which students are not the best source of data. For instance, SETMED-L asks students to evaluate whether the amount of content covered in the course is appropriate. This is an assessment best left to the a peer evaluator with knowledge of the whole curriculum.

Unfortunately, SETMED-L was developed specifically to assess the effectiveness of lectures in medical school, so the items may not be appropriate for UMD. Moreover, the authors investigated the efficacy of SETMED-L at two medical schools and found that

the instrument performed differently at the sites. These finding suggest that UMD should create a SET that is appropriate to the culture and practices at UMD, rather than blindly relying on an instrument created for another school.

The committee could also investigate the administration of SET. Currently,UMDs SET data are collected at the conclusion of a course. However, research suggests that student input may be best solicited during the course, when the instructor can still respond to feedback (Brown, 2008). Research also suggests that providing training to students on the role and importance of SET contributes to validity and improves response rate (Spooren and Christiaens(2017). At present, student training on SET is minimal at UMD.

UMD should investigate the current use of set for personnel decisions and provide training to administrators and faculty on the appropriate interpretation and use of those data. The literature suggests that SET is often used as the sole/primary source of data for making personnel decisions (e.g., promotion, retention, tenure, merit pay) and giving teaching awarrds. Whether or not UMD decides to revise the existing SET, it is important to provide guidance to schools, departments, and faculty on the appropriate way to analyze and use SET data (for an overview of common mistakes, see Hornstein, 2017; Linse, 2017; Stark & Freishtat, 2014). This will help UMD avoid unintentionally disincentivizing effective teaching practices (e.g. active learning techniques or using data to improve instruction; Darwin, 2017, McKeachie, 1997) and mitigate the risk of litigation (e.g., Maffly, 2011). In this effort, a group of educational measurement experts would be very valuable.

Preliminary Assessments of Instrument Functionality, Reliability and Validity (Fall 2006)

In Fall 2006, UM began piloting the University-wide course evaluation items. IRPA's examination of the descriptive statistics for the items revealed highly skewed response distributions; that is, the majority of students used only the positive end of the scale. All 13 Likert-scale items are highly related to each other and to one component, suggesting that the standardized questions are targeting a single topic of "overall" course effectiveness or satisfaction. IRPA's results do not seem to indicate that students view items relating to the course and items relating to the instructor as two distinct aspects of course evaluation.

## Phone Interview Project (Spring 2009)

In spring 2009, IRPA conducted interviews with faculty who had high response rates. It identified anecdotal best practices, such as verbal and electronic reminders, and actions that demonstrate the faculty member's opinion that teaching is important. Interviews with students who did not fill out any evaluations identified that the most popular reason for not participating was that they were too busy and/or ran out of time.

Relationship between Response Rates and Ratings (Fall 2009)

In fall 2009, IRPA found that a visual inspection of average instructor score by response rate bands does not suggest a strong linear relationship. A multiple linear regression analysis showed the relationship between response rate and instructor score, although positive and statistically significant, has little practical significance. It noted that there is a large proportion of the variation in instructor score (95%) that cannot be explained when class size, course level, response rate, and academic discipline are taken into account.

Course Evaluation Differences by Instructor Race/Ethnicity/Citizenship and Gender (Spring 2018)

At the request of the Office of Faculty Affairs, IRPA studied whether differences in course evaluation results can be explained by differences in instructors' race/ethnicity/citizenship and gender. Given there is no "ground truth" measure of instructor quality, the study could not assess potential bias in evaluations. The study found that "there is little evidence for consistent differences between ratings for male and female instructors. Though there is some evidence for differences between race/ethnicity/citizenship categories, these differences are very small." The study did not address open-ended comments, which often inform impressions of bias more than numerical responses.
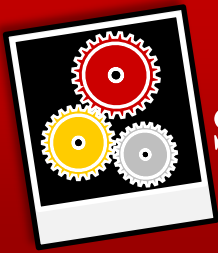
This Campus Assessment Working Group (CAWG) Snapshot reports findings on junior and senior students' perceptions syllabi and how they select courses. The data included represent results from the University of Maryland Student Survey (UMSS), an annual survey administered by the CAWG Assessing Campus Experiences Subgroup (ACES). Respondents complete the survey during the spring semester in Professional Writing courses.

During the spring 2015 semester, out of 3,272 juniors and seniors enrolled in spring semester Professional Writing courses 2,201 (67%) completed the survey.

Race/Ethnicity:
- 53% were White:U.S.
- 16% were Asian:U.S.
- 13% were Black or African American:U.S.
- 8% were Hispanic:U.S.
- 4% were Foreign
- 4% were Two or More Races:U.S.
- 2% were Unknown:U.S.
- <1% were classified as Other, including American Indian and Hawaiian:U.S.

Gender:
- 53% were male
- 47% were female.

GPA:
- 31% had a GPA of 3.50 – 4.00 (the range for which students earn honors)
- 63% had a GPA of 2.30-3.49
- 6% had a GPA of 0.00 – 2.29 (the range for which students are flagged for advising intervention)

The demographic breakdown of respondents is representative of the university as a whole. The data below represent only the responses of survey respondents, not all UMD students; therefore, use caution when generalizing. Percentages may not sum to 100 due to rounding.

Syllabus Resources

In September 2012, the Student Government Association (SGA) passed a bill urging the University Senate to establish a policy to make syllabi available during class registration and the Senate voted in favor. In February 2016, the SGA passed a resolution urging the University to implement the syllabus bill passed earlier. Source: http://www.dbknews.com/2016/02/16/umd-sga-passes-syllabus-resolution-to-revive-university-senate-bill/

The University is committed to finding a solution and is actively working to provide one that meets these requirements. CAWG surveyed students on these questions to understand the value of syllabi and their experiences selecting courses. Note that students responded to these questions based on their current expectations and uses of resources, but these responses may change depending on increased availability of syllabi.
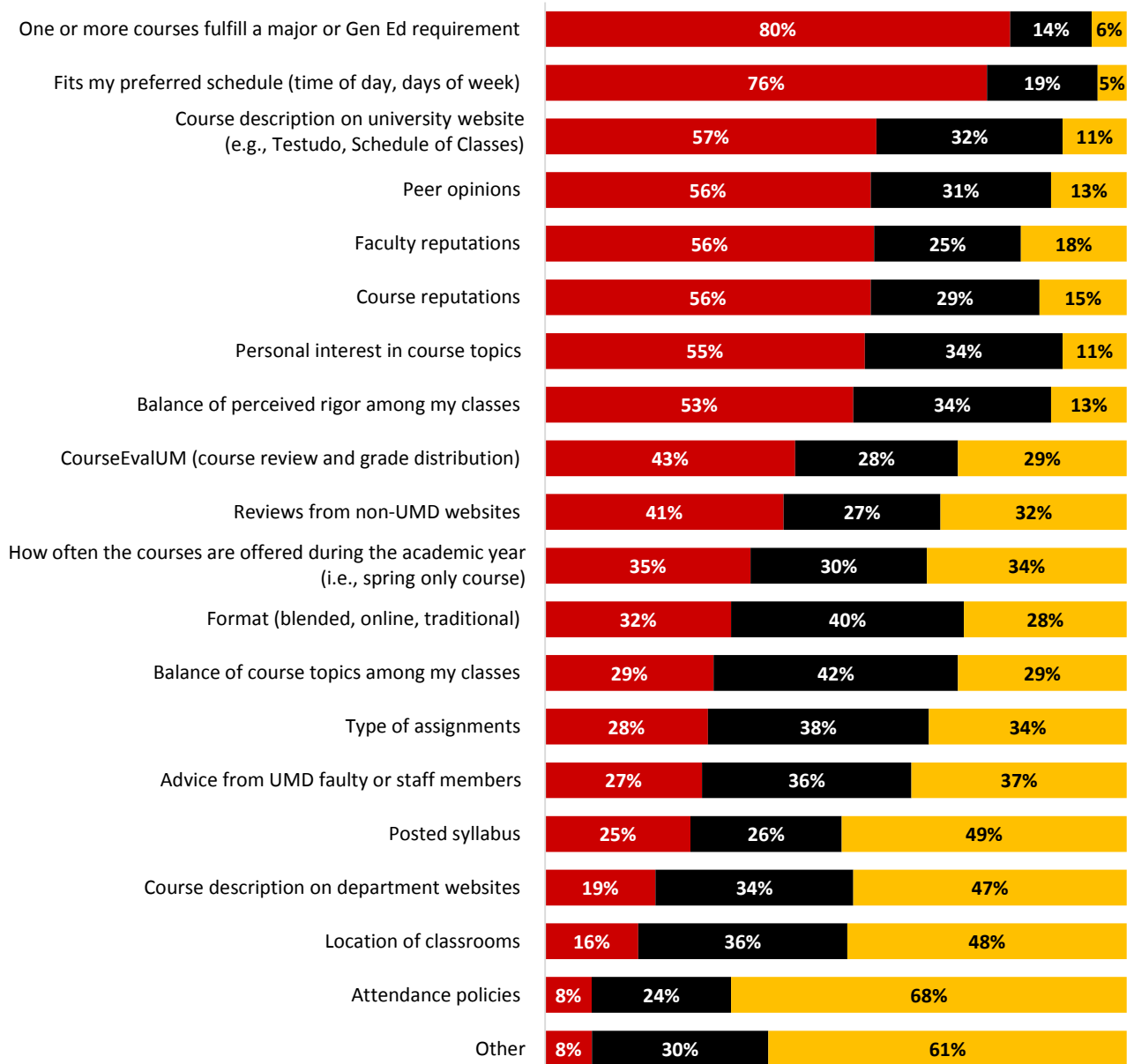
The Faculty Handbook and the Teaching and Learning Transformation Center (TLTC) have resources for creating syllabi:
- Syllabus Guidelines: https://faculty.umd.edu/teach/syllabus.html
- Useful Information for Preparing the Syllabus: https://faculty.umd.edu/teach/useful.html
- Beyond the Guidelines – Writing a Great Syllabus: http://tltc.umd.edu/beyond-guidelines-writing-great-syllabus

Selecting Courses

To what degree did you consider the following factors in choosing your courses this semester?

Legend: ■ A major factor   ■ A minor factor   ■ Not a factor

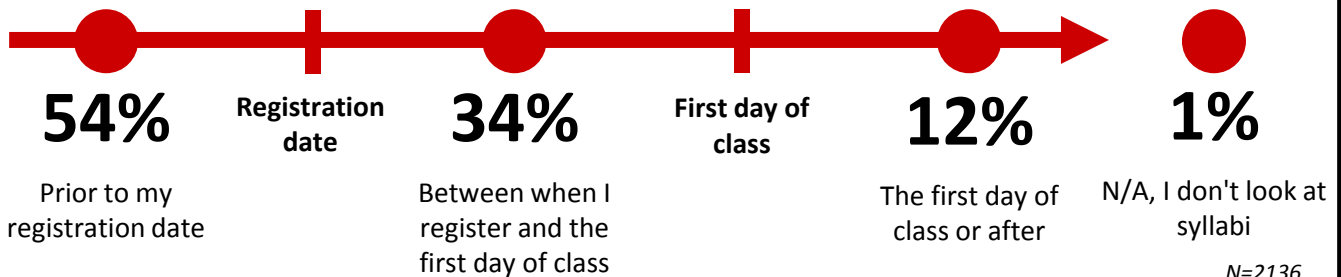| Factor	A major factor	A minor factor	Not a factor
One or more courses fulfill a major or Gen Ed requirement	80%	14%	6%
Fits my preferred schedule (time of day, days of week)	76%	19%	5%
Course description on university website (e.g., Testudo, Schedule of Classes)	57%	32%	11%
Peer opinions	56%	31%	13%
Faculty reputations	56%	25%	18%
Course reputations	56%	29%	15%
Personal interest in course topics	55%	34%	11%
Balance of perceived rigor among my classes	53%	34%	13%
CourseEvalUM (course review and grade distribution)	43%	28%	29%
Reviews from non-UMD websites	41%	27%	32%
How often the courses are offered during the academic year (i.e., spring only course)	35%	30%	34%
Format (blended, online, traditional)	32%	40%	28%
Balance of course topics among my classes	29%	42%	29%
Type of assignments	28%	38%	34%
Advice from UMD faulty or staff members	27%	36%	37%
Posted syllabus	25%	26%	49%
Course description on department websites	19%	34%	47%
Location of classrooms	16%	36%	48%
Attendance policies	8%	24%	68%
Other	8%	30%	61%

N=2183-2196, except "Other" where N=1583

- Logistical considerations are major factors – fulfilling a major or Gen Ed requirement (80%) and fitting a preferred schedule (76%).
- Course descriptions are more of a factor when posted on the university website than when posted on a departmental website (57% compared to 19%).
- Reputations and opinions (perception, faculty reputation, course reputation, reviews and evaluations) are more often a major factor in choosing classes than actual class design (format, posted syllabus, attendance policies, types of assignments).
- Fewer respondents cite location of classes and attendance policy as major factors in choosing courses.

Using Syllabi

Ideally, when would you FIRST like to see a syllabus? Select only one.

54% — Registration date — **34%** — First day of class — **12%** — **1%**

Prior to my registration date

Between when I register and the first day of class

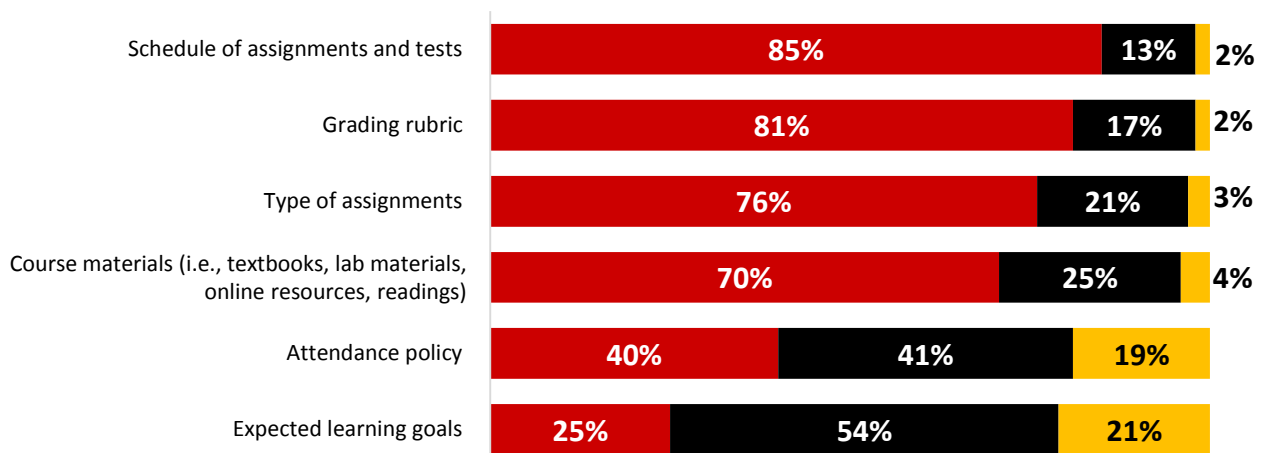The first day of class or after

N/A, I don't look at syllabi

N=2136

- The majority (54%) of respondents would like to see the syllabus prior to registration, and 88% want to see the syllabus prior to the first day of class.

> **Questions to consider:** Why would 54% say that they want to see a syllabus prior to their registration date but only 25% consider a posted syllabus a major factor when choosing courses? How might these numbers change if all syllabi were made available earlier?

How valuable are the following elements of a syllabus?

■ Very valuable   ■ Somewhat valuable   ■ Not at all valuable

Element	Very valuable	Somewhat valuable	Not at all valuable
Schedule of assignments and tests	85%	13%	2%
Grading rubric	81%	17%	2%
Type of assignments	76%	21%	3%
Course materials (i.e., textbooks, lab materials, online resources, readings)	70%	25%	4%
Attendance policy	40%	41%	19%
Expected learning goals	25%	54%	21%

N=2182-2192

- Course design (schedule, grading rubric, assignments, course materials) is very valuable to more respondents than expected learning goals (25%).

Constructs that Assess Baseline and Best Practices in Teaching Effectiveness

- **Timely feedback** (e.g. "I get timely feedback on my work" or "The instructor returned assignments and exams in a timely manner")

- **Clear assignment expectations** (e.g. "Assignment expectations are clear to me" or "The instructor provided guidance for understanding course exercises")

- **Clear grading expectations** (e.g. "Grading criteria are clear to me" or "The instructor grades consistently with the evaluation criteria")

- **Focuses on course content in class sessions** (e.g. "Class sessions help me learn course material" or "The instructor used time effectively")

- **Value of required texts** (e.g. "The required texts (e.g., books, course packs, online resources) help me learn course material")

- **Climate** (e.g. "The instructor helps students feel welcome" or "The instructor treats students with respect")

- **Instructor support** (e.g. "I think the instructor wants students to succeed" or "The instructor was helpful when I had difficulties or questions")

- **Quality feedback** (e.g. "The feedback (e.g., grades, comments, discussions, rubric scores) I get from the instructor helps me improve" or "The instructor provided constructive feedback")

- **Scaffolding** (e.g. "My instructor helps me understand new content by connecting it to things I already know" or "The course presented skills in a helpful sequence")

- **Cognitive engagement and/or rigor** (e.g. "The course developed my ability to think critically about the subject" or "This course was intellectually challenging")

- **Alignment of instruction to assessment** (e.g. "Assessments (e.g., tests, quizzes, papers) relate to course content" or "Graded assignments helped me understand the course material")

Constructs that Inform Student Registration Decisions

- **Course satisfaction** (e.g. "I would recommend this class" or "This course made me want to learn more about the subject")

- **Instructor satisfaction** (e.g. "I would take another course from this instructor if given the opportunity" or "I consistently enjoyed coming to class" or "I enjoyed learning from this instructor")

- **Time invested** (e.g. "On average, about how much time did you spend on this class each week (e.g., doing homework, meeting with project team, studying)?")

- **Major/Non-Major** (e.g. "How does this class fit into your academic plan or course of study?")

Constructs for Open-Ended Feedback

- **Positive aspects** (e.g. "What did the instructor do that helped improve your learning in this course?")

- **Areas for improvement** (e.g. "What could the instructor do better or differently next time to help improve your learning in this course?")
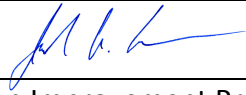
Constructs Related to Teaching Assistants

- **Climate**
- **Timely feedback**
- **Effective use of class time**
- **Open-ended item on positive aspects**
- **Open-ended item on areas for improvement**

University Senate
CHARGE

Date:	February 3, 2017
To:	Philip Evers Chair, Academic Procedures & Standards
From:	Jordan A. Goodman Chair, University Senate
Subject:	Student Course Evaluation Improvement Project
Senate Document #:	16-17-24
Deadline:	December 15, 2017

The Senate Executive Committee (SEC) requests that the Academic Procedures & Standards Committee review the attached proposal that requests a review of the University's student course evaluation system and assess whether changes are needed.

Specifically, we ask that you:

1. Review the report and recommendations of the Task Force on Course Evaluations and Teaching (Senate Doc. No. 02-03-39)

2. Review the Re-evaluation of the Student Teacher Evaluations at UMD (Senate Doc. No. 10-11-06)

3. Review evidence-based best practices regarding student course evaluation systems and procedures at peer institutions and other Big 10 institutions.

4. Consider current scholarship related to course assessment.

5. Consult with various campus stakeholders (e.g., faculty, students, advisors, departmental and college leadership) to better understand their perspectives on current needs, frustrations, and points of satisfaction with the current evaluation process.

6. Consult with a representative from the Teaching and Learning Transformation Center.

7. Consult with a representative of the Office of Institutional Research, Planning, and Assessment (IRPA).

8. Consult with the University's Office of General Counsel on any proposed changes.

9. If appropriate, recommend whether the existing evaluation system including questions and processes related student evaluations should be revised and submit recommended revisions for Senate consideration.

10. If appropriate, recommend an evaluation strategy that utilizes incremental and comparative studies of any necessary changes to the student evaluation system in order to facilitate broad implementation.

We ask that you submit your report and recommendations to the Senate Office no later than December 15, 2017. If you have any questions or need assistance, please contact Reka Montfort in the Senate Office at 301-405-5804 or reka@umd.edu.

Attachment

JAG/rm

University Senate
PROPOSAL FORM

Name:	Benjamin Bederson & Alice Donlan
Date:	January 19, 2017
Title of Proposal:	Student Course Evaluation Improvement Project
Phone Number:	301-405-3394
Email Address:	bederson@umd.edu; adonlan@umd.edu
Campus Address:	4120 McKeldin Library
Unit/Department/College:	Teaching and Learning Transformation Center (TLTC)
Constituency (faculty, staff, undergraduate, graduate):	Faculty, staff
Description of issue/concern/policy in question:	Over ten years ago, the University of Maryland instituted student course evaluations on campus based on work from the May 2004 report to the Senate from the Task Force on Course Evaluations and Teaching (Senate document #02-03-39) which preceded Senate bill 10-11-06. There were 4 primary purposes of these evaluations articulated in the 2005 Final Report: a. Formative evaluation: to provide diagnostic feedback to faculty for the improvement of teaching b. Summative evaluation: to provide one measure of teaching effectiveness for use in the APT and post tenure review processes and in annual productivity reviews c. Informative evaluation: to provide information to students for their use in the selection of courses and instructors d. Outcome evaluation: for the purposes of documenting student learning. The Task Force outlined several recommendations to aid in the pursuit of these four purposes, including a recommendation that the University have a university-wide requirement for student evaluations in all undergraduate and graduate courses. Then, in 2010, the SEC received a proposal requesting a review of the current processes for course evaluations and the APAS Committee was tasked with reviewing the course evaluation system and considering whether it was consistent with the intent of the earlier Senate actions. The resulting Senate bill #10-11-06 recommended a

| | few changes to the course evaluations system, including continued oversight of the CourseEvalUM system by a shared governance body, the development of unit-specific questions, and renewed consideration of a few specific issues, including how to better meet student needs through the course evaluations, how to educate students on the importance of civility in responses, and what efforts need to be made to ensure that APT dossiers include diverse documentation of teaching effectiveness.

While the first instantiation of course evaluations made considerable progress, future efforts can build off of these recommendations to incorporate them into practice. We believe more can be done to improve the content and process of course evaluations to make the process more useful to campus stakeholders.

Three concerns make this proposal particularly timely. First, the current system asks a parallel set of questions for student viewing, and personnel decisions, doubling the length of the survey instead of using questions for multiple-purposes. Second, principal components analysis of current evaluation data has shown that the current questions measure one overarching factor of course satisfaction, as opposed to measuring multiple, theoretically-grounded education constructs as it was originally designed to do. Third, recent research has identified significant bias in most student course evaluations that disadvantage female, ethnic minority, and other groups of instructors. |
|---|---|
| **Description of action/changes you would like to see implemented and why:** | We propose a process to evaluate and revise the current questions and procedures for course evaluations. In particular, we recommend designing the course evaluation to measure four pillars of effective education that comes from the education scholarly literature: classroom climate, course content, teaching practices, and assessment.

● **Classroom Climate:** Is the classroom environment constructed by the instructor inclusive and supportive of learning?
● **Course Content:** Is the content up-to-date, appropriate for the level of the course, and relevant for learners?
● **Teaching Practices:** Does the instructor include evidence-based teaching practices, such as providing timely feedback, scaffolding new information on to prior knowledge, and incorporating active learning assignments?
● **Assessment:** Are the assessments of learning (e.g., tests, quizzes, graded assignments) valid metrics of learning outcomes? |

	Structuring the evaluation around these constructs will more effectively address the four stated purposes of course evaluations. We also anticipate that asking students about concrete classroom activities and practices instead of ambiguous questions about course satisfaction will serve to reduce bias.
Suggestions for how your proposal could be put into practice:	We recommend that the group tasked with addressing this issue perform several activities by first consulting with multiple campus stakeholders (e.g., faculty, departmental and college leadership, students, student leaders, etc.) to understand current needs, frustrations, and points of satisfaction with the current evaluation process. We recommend working closely with the Teaching and Learning Transformation Center (that has performed a preliminary review of other Big 10 school practices and scholarship) as well as IRPA to improve the process of course evaluation. They should also evaluate the best practices of other institutions and the current scholarship on course evaluations. The group should make recommendations to revise the evaluation questions and processes based on what it learns about campus needs and evidence-based best practices. We would suggest that the committee should develop its recommendation through incremental and comparative studies, so that any changes are well understood before being broadly implemented. The University could enact an experimental process that might include, for example, including new and old questions in the same class to compare them directly.
Additional Information:	