

Appendix 1: Past Senate Action on Course Evaluations

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 instructor was responsive to student concerns.
4. The instructor helped create an atmosphere that kept me engaged in course content.
5. Based on the quality of my work in this course, the grades I earned were... (Too Low, Appropriate, Too High)
6. Given the course level and number of credits, the workload was... (Too Low, Appropriate, Too High)
7. 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)

Appendix 3: Number of Evaluation Items by Unit

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	
<i>University-wide</i>	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

Appendix 4: Big 10 and Peer Institution Research on Course Evaluations

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:
 - <http://surveys.indiana.edu/SurveyInstruments/OCQCurrent.pdf>
 - 2012 Report of Item Development & Validity Testing:
http://surveys.iub.edu/OCQ/Developing_OCQ_report.pdf
 - 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 than 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:
 - <http://surveys.iub.edu/OCQ/>
 - White paper on Administering Course Evaluations Online:
http://surveys.iub.edu/OCQ/white_paper%20online_course_evaluations.pdf
 - <http://www.indiana.edu/~bfc/docs/circulars/14-15/B29-2015Revised.pdf>

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: <http://fod.msu.edu/oir/mid-term-student-feedback>
- Tools for evaluating online courses: <http://fod.msu.edu/oir/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/faculty/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:
<https://registrar.osu.edu/faculty/sei/instructorreports.asp>
- Handbook: https://registrar.osu.edu/faculty/sei/SEI_Handbook.pdf
 - Biology Dept opted out of this system. Why?
- Student view: <https://registrar.osu.edu/faculty/sei/studentonlineview.pdf>

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.schreyerstitute.psu.edu/IncreaseSRTERespRate/>
- Faculty FAQ: <https://www.srte.psu.edu/OnlineFAQ/>

Purdue University

- All questions: <https://www.purdue.edu/cie/Website%20CoursEval/courseeval/catalog.pdf>
 - 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."
- Information hub: <https://www.purdue.edu/cie/Website%20CoursEval/courseeval/>

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://sirs.ctaar.rutgers.edu/>
- 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/srt/Images/Student%20Rating%20of%20Teaching%20Form%202015.pdf>
- 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/srt/help/faq.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
- http://ro.umich.edu/evals/CU_TE_TeachingEvals_QR.pdf
- <http://ro.umich.edu/evals/whats-new.php>

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](#).
- SAMPLE paper form: <https://provost.wisc.edu/assessment/documents/100113-Testing%20and%20Evaluation%20Course%20Evaluation%20Example.pdf>
- 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.”

See attachment here:

<https://drive.google.com/a/terpmail.umd.edu/file/d/0B9TPsUDimlyuTmpyd0Y1UkRTV05lQUExXzJCRGxhVnhFbXFZ/view?usp=sharing>

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.”

<https://drive.google.com/a/terpmail.umd.edu/file/d/0B9TPsUDimIyuS0YzYU0yMDM0YWxGLVImRU9YUEdqQ2IPMVYw/view?usp=sharing>

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.”



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 findings 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 awards. 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.

Appendix 6: IRPA Studies on UMD's Course Evaluation System

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.

The report is included below.

**STUDY OF COURSE EVALUATION DIFFERENCES BY INSTRUCTOR
RACE/ETHNICITY/CITIZENSHIP AND GENDER AT THE UNIVERSITY OF MARYLAND**

Amber Bloomfield, IRPA, February 2018

At the request of Faculty Affairs, the Office of Institutional Research, Planning and Assessment investigated the evidence for potential differences in the course evaluations used at University of Maryland. We explored ratings from the five Likert-scale items that are available to administrators (see Appendix) for all fall and spring semester undergraduate courses from Spring 2015 through Spring 2017.

To clarify the intentions and findings of this study it is first important to outline what these analyses are and, just as importantly, are not intended to investigate. These analyses investigate whether there are systematic differences at University of Maryland in the evaluation ratings received by instructors of different genders or race/ethnicity/citizenship categories. The analyses did not directly examine the presence of differences within specific departments; the pattern of course evaluations within a department might differ from the general pattern across the University. Further, this investigation does not explore the validity of the course evaluation items, in that we do not assess their ability to measure actual differences in course or instructor quality. Because we do not have a “ground truth” measure of course or instructor quality, these analyses cannot explore bias in evaluations. Finally, these analyses explore differences only in the five numerical evaluation items available to administrators, not in written comments or responses to other evaluation items. Table 1 shows the means and standard deviations in ratings by instructor gender and Table 2 shows this information by instructor race/ethnicity/citizenship category.

Table 1. Means (standard deviations) for raw administrator item ratings by instructor gender

GENDER	CHALLENGING	LEARNED A LOT	RESPECT	WELL-PREPARED	EFFECTIVE	AVERAGE	N
FEMALE	3.03 (0.96)	3.09 (0.98)	3.51 (0.77)	3.45 (0.82)	3.22 (1.03)	3.21 (0.74)	3,037
MALE	3.19 (0.89)	3.10 (0.98)	3.45 (0.79)	3.40 (0.83)	3.09 (1.09)	3.22 (0.73)	3,576

Table 2. Means (standard deviations) for raw administrator item ratings by instructor race/ethnicity/citizenship category

RACE/ETHNICITY/CITIZENSHIP CATEGORY	CHALLENGING	LEARNED A LOT	RESPECT	WELL-PREPARED	EFFECTIVE	AVERAGE	N
WHITE	3.13 (0.91)	3.11 (0.97)	3.48 (0.78)	3.43 (0.82)	3.16 (1.06)	3.24 (0.71)	3,473
NONWHITE	3.08 (0.96)	3.11 (0.98)	3.49 (0.77)	3.42 (0.83)	3.14 (1.07)	3.22 (0.73)	1,435
INTL	3.25 (0.85)	3.01 (1.01)	3.38 (0.85)	3.34 (0.88)	2.89 (1.20)	3.15 (0.77)	1,011
UNKNOWN	3.06 (0.95)	3.09 (0.98)	3.48 (0.79)	3.42 (0.84)	3.16 (1.06)	3.17 (0.77)	932

Using a mixed effects linear regression model, we examined the relationships between instructor gender and race/ethnicity/citizenship category¹ and average administrator item rating while controlling for other instructor attributes, course attributes, student attributes, and responses to other items on the evaluation instrument (see Appendix for full list of covariates). The data for all instructors of record who received ratings in one or more of the fall and spring terms between Spring 2015 and Spring 2017 were included, regardless of job category. Figure 1 shows the model’s estimated average ratings by instructor gender and race/ethnicity/citizenship category (note that these are predicted ratings and will not match the

¹ We coded “American Indian or Alaska Native,” “Asian,” “Black or African American,” “Hispanic,” “Native Hawaiian or Other Pacific Islander,” and “Two or More” race categories as “Non-White” for analyses. Some of the race categories have very few members, which can complicate analyses; further, previous studies investigating the impact of instructor race on course evaluations have focused on the distinction between White instructors and instructors in other race categories. International instructors are instructors who are not US citizens, identified in our database as “Foreign” based on visa status. Instructors who are naturalized US citizens or permanent residents are not included in this category; these instructors are included in the “White,” “Non-White,” or “Unknown” categories as appropriate.

numbers shown in Tables 1 and 2). The relationship between gender and average ratings was not significant. There was a small but statistically significant impact of race/ethnicity/citizenship category. White instructors received slightly higher ratings across the five items (3.29) than Non-White (3.22), International (3.16), and instructors of unknown race/ethnicity (3.22). The differences appear negligible practically; their statistical significance is likely due to the large number of ratings included in the analysis.



Figure 1. Model-predicted administrator item ratings

Using a second mixed effects linear regression model, we also examined differences in how the individual items behave, as compared to the other administrator items, by instructor gender and race/ethnicity/citizenship category. The two items that show different patterns for instructors of different genders or racial/citizenship categories were the *Challenging* and *Effective* items (Figure 2; note that these are predicted ratings and will not match those shown in Tables 1 and 2). Although both genders tended to receive lower ratings on the *Challenging* item compared to their average rating across the other administrator items, the difference is estimated to be more negative for female instructors (3.00 versus 3.26), than for male instructors (3.19 versus 3.26). In contrast, ratings on the *Effective* item are estimated to be lower than ratings on other items for male instructors (3.12 versus 3.28), but to differ little for female instructors (3.19 versus 3.22). Ratings on the *Challenging* item are estimated to be lower for White, Non-White and instructors of unknown race/ethnicity compared to their ratings on the other administrator items (3.13 versus 3.28, 3.05 versus 3.22, and 3.07 versus 3.25, respectively), while estimated ratings for International instructors are similar on this item (3.22

versus 3.21). Instructors in all race/ethnicity/citizenship categories get lower ratings for the *Effective* item in comparison to the other administrator items (3.17 versus 3.26 for White instructors, 3.12 versus 3.22 for Non-White instructors, and 3.15 versus 3.23 for instructors of unknown race/ethnicity), but the difference for International instructors is larger (3.01 versus 3.26). In all cases, differences in ratings between gender and race/ethnicity/citizenship categories on the individual items are less than a fifth of a point on the 0-4 scale.

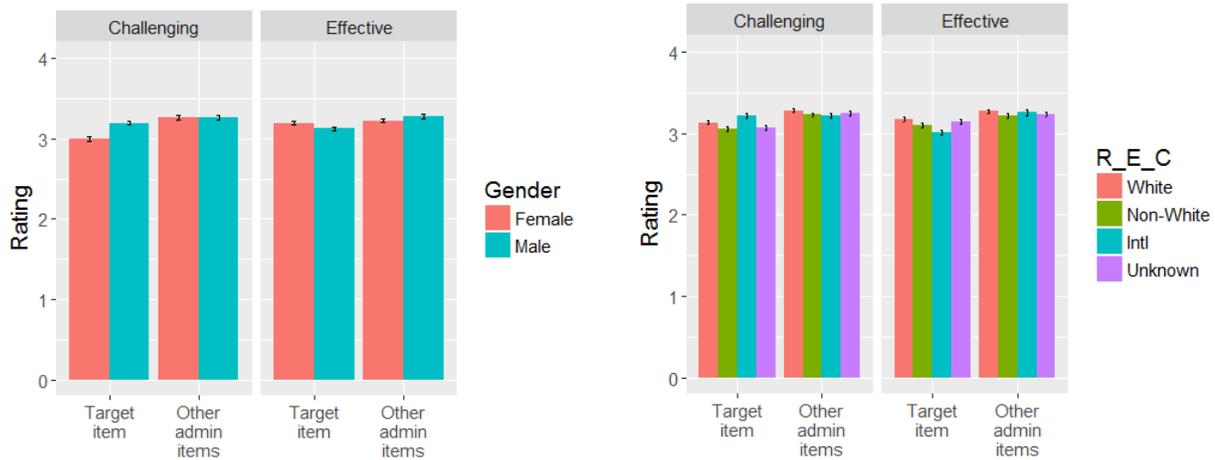


Figure 2. Model-predicted rating on Challenging and Effective items vs. other admin items

Summary

The results of our analyses indicate that, in terms of average administrator item scores, 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 results reported here represent an initial analysis investigating differences in course evaluations based on faculty characteristics.

Appendix

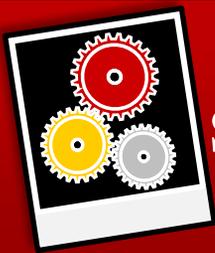
1. Administrator items included as dependent variables
 - a. *Respect*: The instructor treated students with respect.
 - b. *Well-prepared*: The instructor was well-prepared for class.
 - c. *Challenging*: The course was intellectually challenging.
 - d. *Learned a lot*: I learned a lot from this course.
 - e. *Effective*: Overall, this instructor was an effective teacher.

2. Covariates
 - a. **Course variables:**
 - Course college
 - Course level
 - Course meeting time
 - Delivery method of course
 - Average grade in course
 - Course credits
 - Term type
 - Academic year
 - Course duration
 - Section size
 - Response ratio for course instance
 - Classification as: Capstone, Diversity, Experiential Learning, Fundamental Studies, General Education, or Internship

 - b. **Instructor variables:**
 - Age
 - Experience
 - Job category (some collapsing of categories)
 - Percentage of experience at UMD
 - Full-time/part-time
 - Gender
 - Race/ethnicity/citizenship category (White, Non-White, or International)

 - c. **Student variables:**
 - Gender
 - Semester GPA
 - Race/ethnicity/citizenship category (White, Non-White, or International)
 - Grade in course
 - Semester attempted hours
 - Class standing
 - New
 - Interaction between student gender and instructor gender
 - Race/ethnicity/citizenship category match with instructor
 - Primary college match to course college

 - d. **Other Course Evaluation items:**
 - The standards the instructor set for students were... (Too Low, Appropriate, Too High)
 - Based on the quality of my work in this course, the grades I earned were...(Too Low, Appropriate, Too High)
 - Given the course level and number of credits, the workload was...(Too Low, Appropriate, Too High)
 - How much effort did you put into the course? (Little, Moderate, Considerable)
 - How does this course fit into your academic plan or course of study? (CORE or General Education Requirement, Major/Certificate/Minor/Program Requirement, Elective)



CAWG SNAPSHOT OF STUDENT EXPERIENCES

SELECTING COURSES AND USING SYLLABI

University of Maryland

2016 – Issue 1, AUGUST

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.

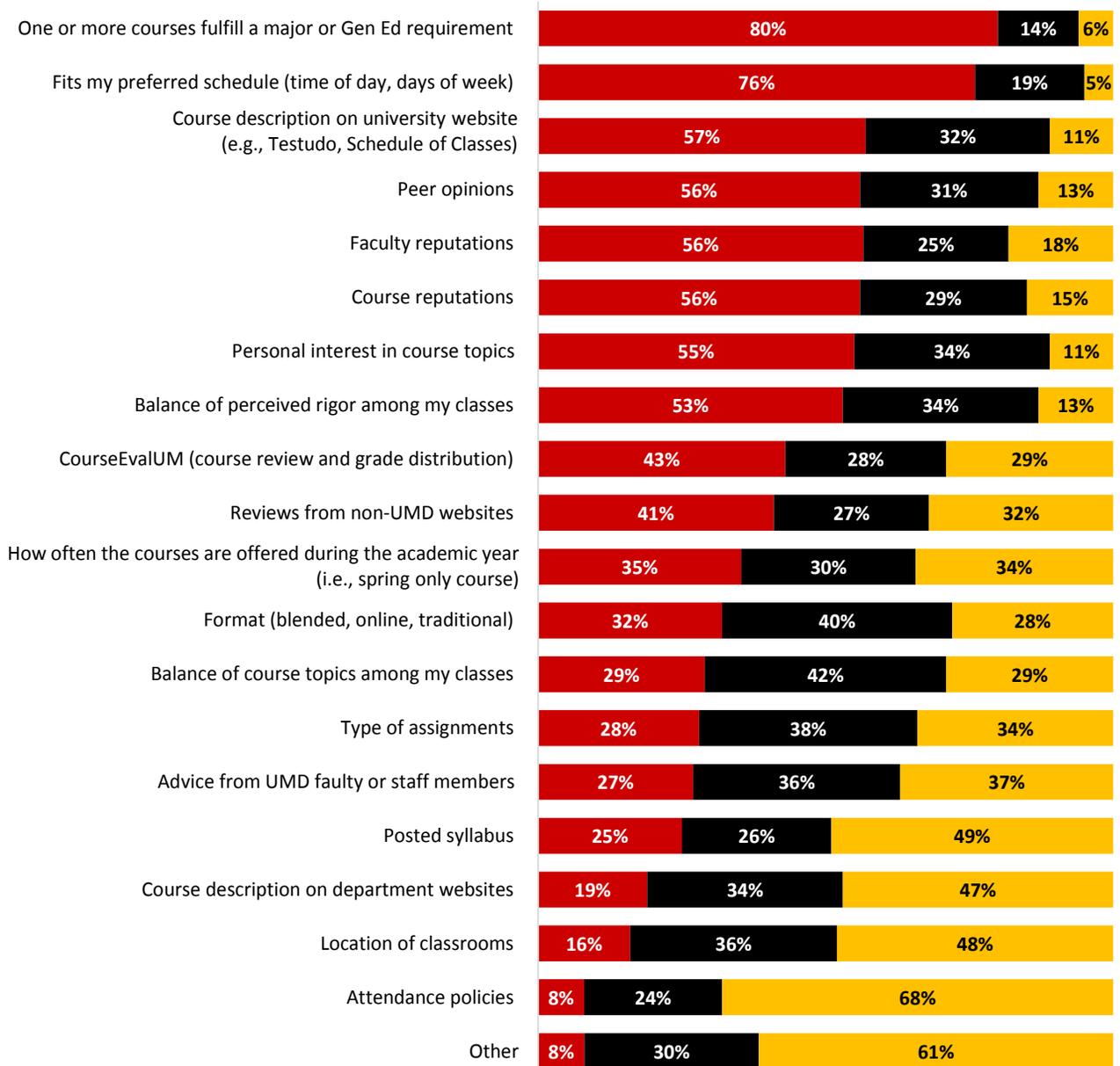
The Faculty Handbook and the Teaching and Learning Transformation Center (TLTC) have resources for creating syllabi:

- 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?

■ A major factor ■ A minor factor ■ Not a factor

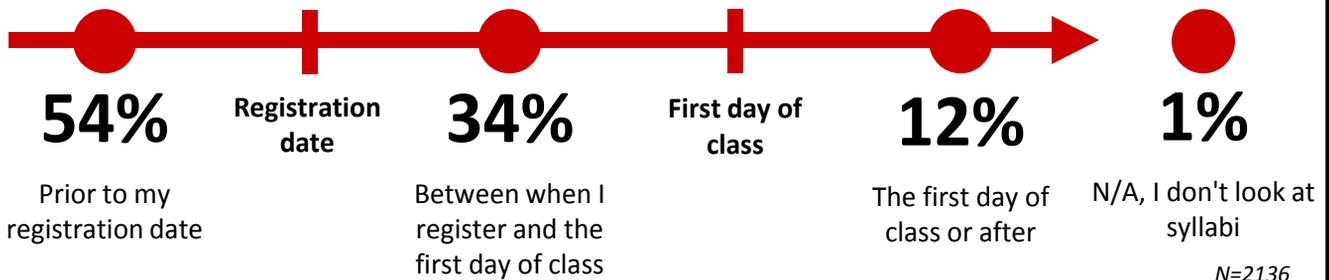


N=2183-2196, except "Other" where N=1583

- 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.



- 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.

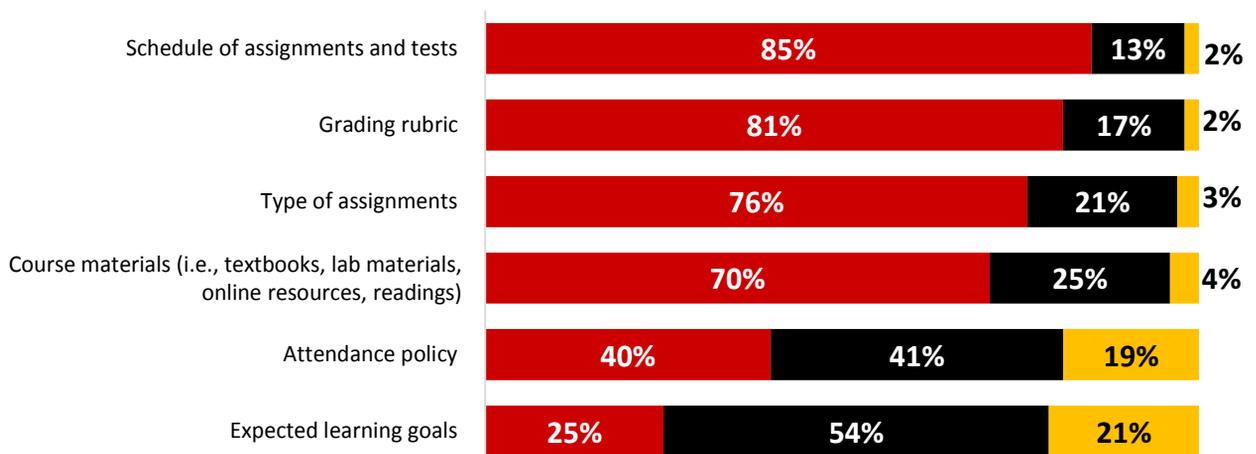
N=2136

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



N=2182-2192

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



The Campus Assessment Working Group (CAWG) regularly gathers and exchanges information about UMD student and alumni experiences. The group is charged with developing a campus "Culture of Evidence" in which data and assessment can inform campus decision making. Its three subgroups focus on freshman experiences, junior/senior student experiences, and retention and completion efforts. For more information, to view past reports, or to join a CAWG subgroup, please visit www.umd.edu/cawg.

Appendix 8: Sample Item Wording for New Constructs

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**

Appendix 9: Charge from the Senate Executive Committee



**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](tel: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:	<p>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:</p> <ol style="list-style-type: none"> 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. <p>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.</p> <p>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</p>

	<p>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.</p> <p>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.</p> <p>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.</p>
<p>Description of action/changes you would like to see implemented and why:</p>	<p>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.</p> <ul style="list-style-type: none"> ● 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?

	<p>Structuring the evaluation around these constructs will more effectively address the four stated purposes of course evaluations.</p> <p>We also anticipate that asking students about concrete classroom activities and practices instead of ambiguous questions about course satisfaction will serve to reduce bias.</p>
<p>Suggestions for how your proposal could be put into practice:</p>	<p>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.</p>
<p>Additional Information:</p>	

Please send your completed form and any supporting documents to senate-admin@umd.edu or University of Maryland Senate Office, 1100 Marie Mount Hall, College Park, MD 20742-7541. Thank you!