

Recommendations for Changes to the Evaluation of Teaching
University of Massachusetts Amherst
Fall 2017

Teaching Evaluation Working Group

Joye Bowman, History

David Ford, Chemical Engineering

David Gross, Biochemistry and Molecular Biology

John Kingston, Linguistics

Jennifer Normanly, Biochemistry and Molecular Biology

Jennifer Randal, Educational Policy, Research, and Administration

Jeffrey Starns, Psychological and Brain Sciences

Martha Stassen, Co-Chair, Office of Academic Planning and Assessment

Gabriela Weaver, Co-Chair, Teaching Excellence and Faculty Development

Recommendations for Changes to the Evaluation of Teaching at UMass

Executive Summary

In order to provide more complete descriptions of faculty teaching and, thereby, more robust approaches to evaluating that teaching, a faculty Working Group was formed to study such approaches and provide suggestions that could be applicable to the UMass campus. More robust approaches are those that will be less sensitive to factors that are not related to student success and more consistent among those factors that are expected as part of the teaching practice within departments, colleges, and across the campus. To this end, this report includes two main sections of review and recommendations. The first section involves changes to the Student Response to Instruction (SRTI) instrument and its implementation, including clarifications about appropriate uses of SRTI statistics. The second section describes a new approach to the evaluation of teaching, involving a customizable framework that uses multiple types and sources of data. This framework is based on research and its use is modeled on similar implementations taking place at other public research institutions (including Association of American Universities institutions). The framework allows for customization, yet results in numerical data that can be compared across contexts, such as different colleges or departments. This framework can also serve as a formative mentoring tool.

Before this document moves forward as a finalized recommendation from the Working Group, we intend for it to be reviewed and commented on by university leadership as well as representatives from the Massachusetts Society of Professors (MSP) union and the Faculty Senate.

Motivation for the Work

The economic prosperity of our nation relies on a talented workforce with the ability to nimbly address new challenges and develop state-of-the-art technologies. Innovative, student-centered teaching practices can promote these skills, as reported in numerous scholarly publications and summarized in agency reports (such as Kinzie and Kuh, 2017; Kober, 2015; National Academies of Sciences, Engineering and Medicine, 2017). Public higher education institutions are in a particularly good position to influence the development of the workforce in this way because the students attending these institutions represent a large and diverse segment of the college-going population.

UMass Amherst has already demonstrated its commitment to high quality teaching by investing in and promoting infrastructure that supports it. For example, approximately six years ago, the campus developed two teaching spaces to be used, on a pilot basis, for collaborative learning

(where students work in teams during class sessions). These spaces were modeled after the MIT “TEAL” classrooms (Massachusetts Institute of Technology, 2017). In combination with this development of learning spaces and under the direction of the Provost, the then-Center for Teaching and Faculty Development began to implement faculty development programs specifically aimed at preparing faculty to teach using evidence-based, student-centered approaches.

The institution further extended its investment in the use of evidence-based educational practices when it approved the construction of an integrated classroom building that would be home to a variety of innovative teaching spaces. Featured in this building are Team-Based Learning (TBL) spaces that include not only furnishings to support group work, but also laptops, flat panel screens, whiteboards with cameras, and a central podium with control center for every group. Faculty who are interested in teaching in these spaces must first go through one of the training series offered by the Institute for Teaching Excellence and Faculty Development (TEFD, formerly Center for Teaching and Faculty Development) to learn about evidence-based teaching practices and best practices in course (re)design.

In spite of these campus investments and the high value placed on student success and teaching in the campus Strategic Plan (UMass Amherst, 2013), there remains a disconnect between the evaluation of teaching and the goal of increasing student success through changes in teaching. In order to improve the undergraduate experience and increase the quality of teaching, campus units are working actively to encourage best practices in pedagogy (through programming offered by units such as TEFD and Instructional Innovation as well as within some colleges and departments). However, traditional methods for the evaluation of teaching may not be sensitive or robust enough to adequately inform this work or fully represent pedagogical successes. For example, when faculty initially began using the UMass Amherst pilot TBL classrooms, faculty response to using these spaces was very positive; however, a preliminary exploratory analysis of SRTI ratings from 30 early TBL courses showed a decrease in scores for a number of instructors in the initial TBL implementation. These scores rebounded in later iterations for 18 of these courses (Office of Academic Planning & Assessment, 2014). This dip is a consequence of early implementations of research-based, but relatively unfamiliar, pedagogical practices. While these pedagogies may be backed by evidence, indicating that they have positive effects on student success, the risk of affecting student response scores may be perceived as too large when these scores have an outsized influence on professional review processes for faculty. In addition, through faculty focus groups and many other conversations, it has come up on numerous occasions that faculty and campus leaders would like to see alternatives or enhancements to teaching evaluation. The belief that teaching is undervalued is widespread at 4-year institutions nationally (Hutchings, Huber, & Ciccone, 2011). Whereas universities rely on a rich set of discipline-sensitive metrics and external reviews to evaluate

faculty research, the evaluation of teaching relies heavily on student ratings (Berk, 2005; Miller & Seldin, 2014; Seldin, 1998; Vasey & Carroll, 2016), often summarized by a single question of overall judgement of the course or instructor. While there are 14 Likert-scale items on the UMass SRTI survey asking about a variety of aspects of the course and its instruction, there are three overarching items: how much students feel they have learned in the course, their overall rating of this instructor's teaching, and their overall rating of the course. The interpretation and use of SRTI results varies substantially by department. However, our Working Group discussions and conversations with Heads, Chairs, and faculty across campus suggest that many departments focus on these summative items for promotion and merit-based evaluations of faculty teaching. In some cases, the overall rating of instruction items makes up the majority of the evaluative record regarding a faculty member's teaching effectiveness. Yet, the correlation of this instrument to other measures of teaching effectiveness (concurrent validity) or objective measures of student learning has not been demonstrated, in part because the campus does not currently use any other campus-wide, systematic measures of teaching effectiveness.

Universities have long relied solely on student surveys as the primary means of evaluating instruction, despite numerous known limitations to this practice. Over 50 years of research into the validity and reliability of student ratings of instruction provide important guidance for survey design and implementation, as well as identification of the factors that can affect the usefulness of student rating results. The ease of collecting student ratings and their applicability across disciplines make them appealing, and this approach provides the most systematic and inclusive method for capturing students' voices on their learning experiences. However, as has been true for decades, there is also a vigorous debate on the validity and reliability of student ratings as measures of effective teaching. The concerns over the use of student ratings are particularly potent when these surveys are the only representation of teaching effectiveness and are used for personnel decisions (Basow and Martin, 2012; Benton, 2012; Spooren, Brockx, & Mortelmans, 2013).

The University's SRTI form was designed based on the recommended practices and concepts for student course rating surveys and follows the professional standards of survey research design. While analyses of the UMass SRTI results have shown no large and systematic effect of instructor gender or race on SRTI scores (Office of Academic Planning & Assessment, May 2016), other literature suggests that student course surveys can be discriminatory with respect to age, gender, and race. As noted in Basow and Martin (2012):

Although there is still a considerable amount of research needed to understand all the ways that student evaluations can be biased, this chapter suggests that not only is some bias possible but it is likely. As a human activity reliant upon person perception and interpersonal judgment, student ratings are affected by the same factors that can potentially affect any rater's

judgment: stereotypes based on gender, race/ethnicity, age, and other qualities (such as professor sexual orientation); the equation of “what is beautiful is good;” more positive feelings towards those who seem to reward us (e.g., with good grades). Even though the size of individual effects may be small, for specific professors these small effects may add up to make a meaningful difference on the ratings they receive. (p. 46).

While still relatively limited, there is also emerging research into the relationship between higher student survey scores and actual student learning. This type of research is challenging to carry out because of the complexities of measuring student learning across multiple instructional environments and linking these individual student performance results to student ratings. Nonetheless, the existing research suggests that higher student ratings are related to grading leniency, lower levels of learning, and poorer student performance in follow-on courses (Stroebe, 2016; Uttl, White, & Gonzalez, 2016). Findings like these suggest that student ratings may actually work *against* high quality teaching practices that challenge students in new ways.

Taken together, research findings about discriminatory response biases and the sacrifice of quality for higher ratings show a complementarity of these limitations that may amplify when underrepresented faculty try to engage in novel teaching practices. These limitations in student ratings suggest that they should, at a minimum, be part of a set of *multiple* measures, as is the practice when evaluating faculty research. This suggestion is consistent with the current guidelines for use of SRTI results (see <http://www.umass.edu/oapa/srti/perform.php>) and recommendations for best practices for student ratings more generally (Benton & Cashin, 2012).

It may be possible to support faculty who are engaged in varied pedagogical methods, including innovative ones, by utilizing approaches to the evaluation of teaching that consider more dimensions of teaching work than what takes place in the classroom. By introducing a scholarly framework for the evaluation of teaching, this proposed approach will help faculty members create a shared vision of good teaching in their discipline, identify appropriate forms of evidence of good teaching, and apply the resulting framework for such means as mentoring, annual evaluations, promotion, and tenure. Such a practice would also be consistent with the spirit and the letter of the UMass MSP union contract, which states:

Article 33.4 Over time, the annual evaluations of teaching should attempt to capture the total contribution of the candidate to the instructional mission, both inside and/or outside the classroom, through multiple modes of evaluation, not just student evaluations.

Further, the Red Book provides that the basic file for major personnel actions (RPT) must include “evaluations of teaching effectiveness, including but not limited to those of students.”

The Work of the Task Force

Weaver and Stassen jointly proposed to the Provost a Working Group to address the need for a better, multi-faceted method for evaluating the quality of teaching. The group would ultimately develop a set of recommendations for ways to improve and expand our tools and practices for teaching evaluation. We noted at the start of our work that there appears to be no single definition for the “quality” of teaching, a challenge when working to identify indicators of instructional quality. However, work and action on better forms of evaluation are beginning to emerge, including from national-level organizations, such as the Association of American Universities, National Research Council, American Association of Colleges & Universities, and the Howard Hughes Medical Institute/Cottrell Scholars. The conversations of the Working Group drew from emerging definitions of quality as well as our own and our departments’ perspectives on quality.

The Working Group was comprised of seven members plus the two Co-Chairs (Weaver, Teaching Effectiveness and Faculty Development (TEFD); Stassen, Office of Academic Planning & Assessment (OAPA)). Members represented a variety of disciplines and colleges and brought with them various levels of expertise to contribute to the multi-faceted discussion of evaluation of teaching at UMass. The Working Group met eight times during the Spring 2016 semester to carry out the following:

1. Review research and policy literature from across the country on teaching evaluation.
2. Review practices and history at this campus.
3. Discuss methods that will be:
 - valid and reliable across different approaches to teaching and different disciplines.
 - feasible for departments/faculty to incorporate, but also usable by administration across disciplines.

The topics and literature for review and discussion included:

- National-level documents on teaching evaluation.
- SRTI and other current practices at UMass.
- Examples of recent change efforts in evaluation approaches at other universities and UMass.
- Methods reviewed:
 - student surveys
 - peer evaluation approaches
 - observation approaches
 - faculty portfolios and faculty reporting

Principles for Evaluation of Teaching Practices at UMass Amherst

Following from our review of current practices and our own experiences, the Working Group identified a set of principles to guide future teaching evaluation activity:

- **Evaluation should include multiple dimensions** of teaching: categories of activities that capture the teaching endeavor in its totality, including aspects that take place outside of the classroom.
- **Evaluation should include multiple lenses:** multiple *sources* and *types* of data, including faculty self-report (e.g., course materials, evidence of student learning and reflections on it), peer input (e.g., class visits, review of course materials, discussions with the instructor), and student voices (e.g., course evaluations, alumni feedback).
- **Evaluation should involve triangulation:** no measure should be used in isolation, and analysis and interpretation should include an acknowledgement of the ways in which these measures provide reinforcing and/or conflicting perspectives on an instructor's effectiveness.
- **Both formative and summative** uses of the data must be possible to maximize the impact on teaching effectiveness. In addition, the evidence should be useful in a longitudinal view (over courses, semesters, and years) so that improvement over time can be documented.
- There must be a **balance between uniformity** across departments **and customization** to different disciplines in order to maximize usefulness to the administration as well as faculty.

Clearly, these principles require us to consider measures of teaching effectiveness beyond the current SRTI student ratings. Drawing from practices at other universities and a review by the American Association for the Advancement of Science (2013), we identified five types of measures currently in use within higher education institutions: Surveys, Interviews, Observations, Teaching Portfolios, and Document Analysis. A description of each of these and a discussion of the strengths and challenges of each is outlined in Appendix A. On our own campus, some of these tools are also being used or piloted. The following is a description of those currently in use at UMass Amherst.

Practices in Evaluation of Teaching Currently Used at UMass

Student Response to Instruction (SRTI) End-of-Semester Surveys

UMass has utilized the SRTI student survey as the only universally applied form of course evaluation across the campus for over 20 years. First adopted in 1995, SRTI was developed in a collaborative effort between OAPA, which administers the survey, the then-Center for Teaching, and the then-Faculty Senate Council on Teaching, Learning, and Instructional Technology. SRTI was initiated by the Council as a replacement for an older campus-wide

student survey instrument that was deemed so unsatisfactory that most departments had abandoned it in favor of implementing their own student survey forms and methods. SRTI was designed to be short, focused on elements of teaching effectiveness that students are capable of judging, and to provide both formative information (Items 1-9, 13) and summative evidence (Items 10-12) to support instructional improvement as well as promotion and merit decisions. Through OAPA, the University administration covered the costs of the survey forms, scanning, and analysis for departments. Participation with SRTI began with a small group of pilot departments, but over time grew to virtually 100% volunteer departmental participation.

Over time, OAPA has made adaptations to the form based on instructor feedback, identification of additional institutional needs, and recommendations from the scholarship on student ratings of instruction. The instrument (attached as Appendix B) shows the student-relevant indicators of teaching effectiveness most commonly identified in studies of student ratings (see, for example, the summary reviews of Berk, 2006 and Hativa, 2013):

- Organization (SRTI Items 1, 4)
- Clarity (SRTI Items 2, 3)
- Instructor engagement/enthusiasm/interest/expressiveness (SRTI Item 5)
- Rapport (SRTI Item 6)

The instrument also addresses other constructs that are often mentioned, but not consistently, across studies:

- Interaction/Questioning (SRTI Item 9)
- Exams/Grades/Student Evaluation (including Feedback) (SRTI Items 7, 8)
- Students' Perceived Learning (SRTI Item 10)
- Workload/Difficulty (SRTI Item 13 – Hours spent working on the course)

SRTI's content validity is further suggested by the fact that the survey items align with UMass Amherst students' description of the characteristics of good General Education courses in a 2007 internal study (see "What students value in General Education courses"

http://www.umass.edu/oapa/oapa/topics/gen_ed_assessment.php):

- Quality of the professor/student relationship (instructor responsiveness to student needs, positive interpersonal climate in the course) (SRTI Items 3, 6, 7)
- Promotion of student engagement (through classroom management strategies like active learning and in subject matter) (SRTI Items 5, 9)
- Course planning (organized, being prepared, clear expectations, assignments) (SRTI Items 1, 2, 4, 8)

The campus is currently making the move from a paper administration of SRTI to an online administration, with full online participation beginning Fall 2017.

Mid-term Assessment Process (MAP)

TEFD has been administering the MAP program for well over a decade. This is a voluntary program, requested by individual faculty members, in which TEFD staff members collect feedback from students in a class during weeks five through nine of the semester. TEFD staff then analyze these data and have a private consultation with the faculty member to discuss the feedback and provide suggestions. The process is intended to be purely formative, and it is kept completely confidential—TEFD shares the data only with the specific faculty member who requests the MAP.

Feedback collected from students consists of two types. In large courses (enrollment larger than 50), students individually answer a Likert-scale survey along with two short-answer questions. In smaller courses, students are asked to form small groups (3-5 students) and discuss four short-answer questions. They develop collective responses to these questions and submit one answer sheet per group. TEFD has been exploring digital methods to allow us to collect and analyze additional qualitative (short-answer) data from larger classes, but this has taken place only on a pilot scale to date. More pilot testing is planned.

College of Natural Sciences (CNS): Dissemination of Best Practices Guidelines for the Reappointment, Tenure, and Promotion Review

CNS has developed a comprehensive set of recommendations for Personnel Committees (PCs) that includes suggestions for teaching evaluation (see Appendix C). These recommendations include guidance in how to report and interpret SRTI results and suggestions for additional evidence that should be brought to bear:

Beyond these ratings, various other evidence of teaching strength such as signed letters from students who have taken classes with the candidate, letters from undergraduate or graduate research advisees, letters from colleagues who have co-taught courses or sat in on classes, evidence for participation in innovative teaching programs, and documentation related to major initiatives in course re-design or pedagogical advances should be included. Letters from students, in particular, help with assessment of the SRTI scores. Also helpful to the CNS PC are comments about how the SRTI scores for the faculty member under review compare with the mean scores obtained by previous instructors who have taught the same course. As noted above, personal statements from the candidate that discuss aspects of teaching excellence are essential. (Appendix C: College of Natural Science Personnel Committee: Best Practices, p. 9).

College of Humanities and Fine Arts (HFA): Peer-to-Peer Teaching Review

HFA has begun a pilot effort to connect faculty members with each other to build constructive conversations about teaching (see <https://www.umass.edu/hfa/peer-peer-teaching-review>). Included in this process is the opportunity for peer observation of teaching, the results of which can provide guidance and support for the instructor's teaching improvement over time. (See Appendix D for the Instructor and the Reviewer Forms used as part of this process.)

Recommendations

In addition to the set of organizing principles we outlined at the start of this report, the Working Group has identified a number of recommended next steps that fall into three areas: SRTI Changes, Recommendation for Generalizable and Customizable Rubric for Evaluation of Teaching, and Consideration of Other Forms of Evidence.

Recommendations Related to SRTI

The issue of how to effectively incorporate student perspectives into the evaluation of teaching was a theme that ran through the Working Group's discussions. To further inform these discussions, OAPA conducted focused one-on-one interviews about the SRTI form with seven faculty members during the Summer of 2016. These individuals included representatives from relevant Faculty Senate Councils, the Rules Committee, the MSP, and members of the Working Group. There was a consensus between both the Working Group discussions and these in-depth interviews that the student perspective is an essential component to a campus system for evaluating instructional effectiveness. In addition, the student perspective should focus only on those aspects of teaching that students are qualified to address, and it should not represent the primary (or only) perspective or source of information used to evaluate teaching effectiveness.

The discussions also raised the following questions and concerns:

- The current SRTI form does not adequately address contemporary teaching practices, including the focus on active learning strategies and campus priorities for innovative instruction.
- On a related point, the current form is very "instructor-centric." There should be more of an emphasis on the student's role in facilitating an effective learning experience.
- Course evaluations may reflect students' biases (including gender, race/ethnicity, age, etc.) and the results may have differential impacts on instructors based on gender, race/ethnicity, age, or other characteristics not related to actual teaching effectiveness.

- Questions about the extent to which SRTI adequately addresses student learning, and some question as to whether the form *should* address student learning (i.e., are students very good judges of what they have learned?).
- Questions about using SRTI for both formative and summative purposes: should these two purposes be separated from each other (either in terms of two separate processes, or separating the formative results—Items 1-9—from the summative results—Items 10-12—and only reporting the summative results for promotion and tenure purposes)?

Specific recommendations fall into four categories:

1. Adapt/Change the Campus’s SRTI Instrument

Most of the conversations affirmed the general appropriateness of the current SRTI items. While participants felt that the foundational elements were appropriate and a wholesale change was not required, they did identify areas where adaptation is required.

SRTI Adaptations

Edit and re-order some of the current items to improve clarity and guard against triggering potential student biases:

1. Change the wording of Item 6 from “The instructor shows a personal interest in helping students learn” to “The instructor shows an interest in helping students learn”
2. Change the order of the items so all items about the instructor are asked together and items about the course are asked together.

(NOTE: These changes will be made in Fall 2017 when the online SRTI system has full campus implementation.)

Additional SRTI Questions and Survey Features

1. Add an item focused on student motivation, interest, effort in the class (tied to a focus in reinforcing to students *their* role/responsibility for their learning).
2. Add an item on clarity of course objectives and purpose of the course (to better focus on student learning).
3. There is less consensus on whether to add an item that focuses on managing diversity and/or being responsive to diversity. Some felt this was an important component to include, while others raised concerns that the item could be potentially misleading or have a chilling effect on instruction.
4. Provide opportunity for tailored questions by faculty or departments, and also

batteries of potential items for TBL and other pedagogical and active learning innovations.

While the group generally accepted the current framework for SRTI, there was also a recurring discussion that emphasized the importance of focusing on student learning, the extent to which faculty course design fosters student learning, and students' responsibility for taking an active role in their learning. This approach was informed by the example of the Student Assessment of Learning Gains (SALG), which asks students to report the extent to which they made gains on a host of learning objectives as a result of their course and the extent to which course features and faculty interactions facilitated their learning (<http://www.salgsite.org/>). While not using the SALG form, Stanford University, Yale University, and Amherst College have all recently added items to incorporate this perspective into their forms.

This approach relies on the articulation of course-based learning objectives (or the use of a generic course objective question, as suggested in addition #2 above), the identification of teaching strategies of which students rate the effectiveness, and a survey and reporting system flexible enough to populate course surveys with course-specific objectives and teaching strategy details.

(NOTE: The new online system will have this capacity for those individual instructors and departments that want to focus on students' experiences tied to specific learning outcomes.)

2. Improve Support and Communication to Departments and Personnel Committees

Discussion here included offering orientation sessions for Heads, Chairs, and Personnel Committees, generating easily accessible and well-publicized documentation of the appropriate uses of SRTI data, and greater guidance for individual instructors in how to effectively use their SRTI results.

3. Improve Reporting of Results

1. Include a statistically appropriate "measure of uncertainty" to aid departments and Personnel Committees in identifying the extent to which observed differences are real and worthy of attention.
2. Consider separate reporting of formative and summative items, with individual faculty receiving a full report of all results, and departments receiving results only for the summative items.

(NOTE: With the help of Jeffrey Starns, Associate Professor, Psychological and Brain Sciences, member of the Working Group, and OAPA Faculty Fellow, changes to the manner in which the results are reported will be made in Summer 2017 with full implementation in Fall 2017.)

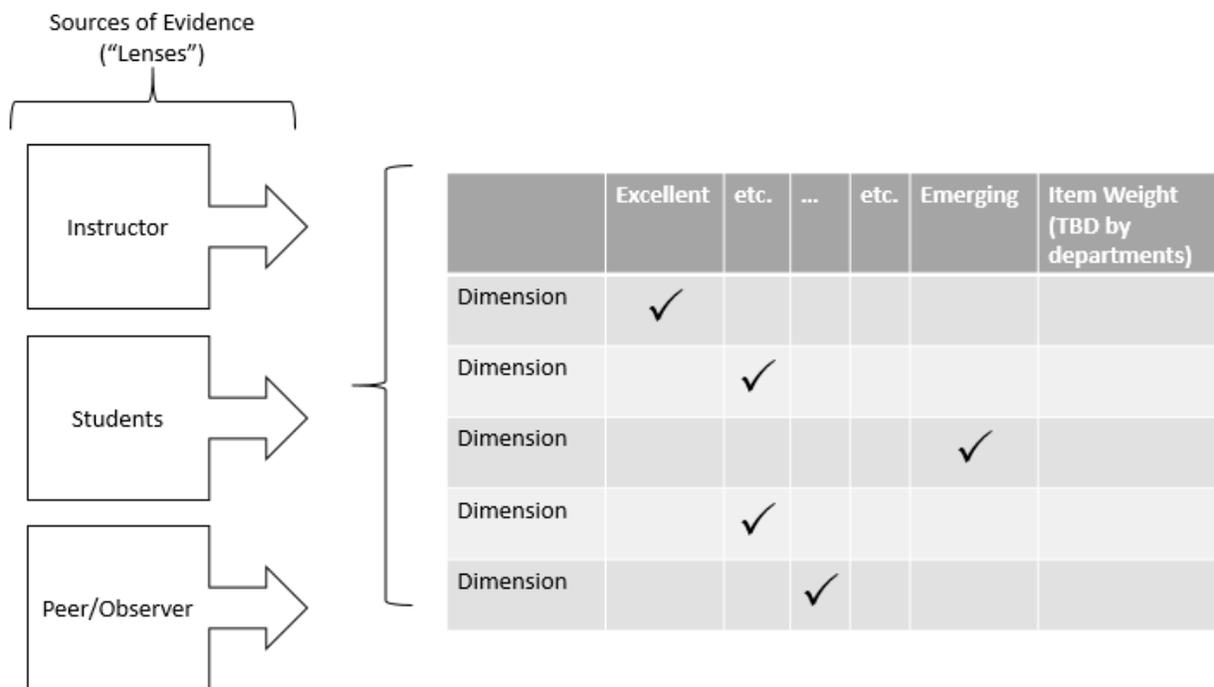
4. Continue and Enhance Analytic Investigations into:

1. Differences by instructor demographics

2. Online administration effects
3. Pedagogical Innovation effects
4. Instrument Concurrent Validity and relationship to student learning/performance.

Recommendation for Generalizable and Customizable Rubric for Evaluation of Teaching

In order to accommodate multiple measures of instructional effectiveness, we propose adapting a tool that will allow for organizing, analyzing, and summarizing results into a comprehensive whole. The framework of a rubric for managing and representing the data is summarized in this figure:



Each "dimension" represents a different aspect of teaching at UMass. Each department would have their own criteria for and ways of defining different levels of achievement within each dimension. Furthermore, a dimension would be evaluated based on evidence provided by multiple sources, representing different lenses onto that dimension. And evidence would emerge from multiple different types of data: survey, observation, reflection, etc.

The Center for Teaching Excellence at the University of Kansas has applied the framework generalized in the above figure and developed the Teaching Effectiveness Scoring Rubric (see Table 1), drawing on the literatures on peer review of teaching and teaching as scholarly inquiry (e.g., Bernstein, 2008; Glassick et al., 1997). This rubric identifies seven dimensions of teaching

practice: Goals, Content, and Alignment; Teaching Practices; Achievement of Learning Outcomes; Classroom Culture and Student Perceptions; Mentoring and Advising; Reflection and Iterative Growth; and Involvement in Teaching Service, Scholarship, or Community. Attention to these dimensions generates a more comprehensive view of the faculty member's teaching contributions, not just the instructor's behavior in the classroom. The rubric provides both guiding questions and defined expectations for each category.

The rubric is intended to be used by department committees to integrate information from **the faculty member** (e.g., course materials, evidence of student learning and reflections on it), **peers** (e.g., class visits, review of course materials, discussions with the faculty member), and **students** (e.g., course evaluations, letters from alumni). The rubric can also be used to guide a constructive peer review process that prompts reflection and iterative improvement by the instructor. The questions and criteria are to be **defined by each department**, and departments should determine how to weight each category and what forms of evidence will be used for each—triangulating with **multiple forms** and **multiple sources** of data for each entry in the rubric.

The rubric is intended to improve evaluation by increasing consistency across reviewers and over time, fostering integration of multiple sources of information, providing structure for a narrative, and generating a numerical yet more nuanced representation of teaching effectiveness. By providing criteria with descriptions of different levels of performance (for example, Below Expectations, Meets Expectations, Exceeds Expectations), the rubric can simultaneously inform evaluation and facilitate developmental scaffolding to guide faculty improvement in the quality of teaching.

In addition to shaping the evaluation process itself, the rubric can be used to articulate a shared vision within departments of effective teaching, one that is aligned with what both discipline-based and more general research says about teaching effectiveness. Importantly, the rubric is designed to be a starting point for department-level conversations about teaching effectiveness. The rubric dimensions are to be held constant across the campus, but departments are invited to adapt the language, interpretation, and relative weights to fit disciplinary expectations, and to use the rubric to build consensus about what effective teaching looks like. This approach provides the campus with a common framework while preserving disciplinary identity and specificity.

Table 1. Teaching Evaluation Rubric, adapted from that developed by the University of Kansas. All of the bulleted items defining the intersections of rows and columns are to be determined at the departmental level.

Dimensions of Teaching	Below Expectations: 1 - 2 <i>Poor (1): Consistently at this level</i> <i>Marginal (2): Some teaching at this level</i>	Meets Expectations: 3 <i>Competent</i>	Exceeds Expectations: 4 - 5 <i>Professional (4): Some teaching at this level</i> <i>Advanced (5): Consistently at this level</i>
Goals, content, and alignment <i>What are students expected to learn from the courses taught? Are course goals appropriately challenging? Are assessments aligned with the learning goals?</i>	<ul style="list-style-type: none"> • Course goals are unclear, inappropriate, or marginally related to curriculum • Content and materials are outdated or unsuitable for students in the courses • Range of course topics is too narrow or too broad • Assessments not clearly aligned with learning goals or institutional expectations 	<ul style="list-style-type: none"> • Course goals are articulated and appropriate for curriculum • Content is current and appropriate for topic, students, and curriculum • Course topics include an appropriate range • Standard, intellectually sound materials 	<ul style="list-style-type: none"> • Course goals are well-articulated, high quality, and clearly connected to program or curricular goals • Content is challenging and innovative or related to current issues and developments in field • Course topics include appropriate range and depth, with integration across topics • Materials are high quality and well-aligned with content and course goals
Teaching practices <i>How is in-class and out-of-class time used? What assignments, assessments, and learning activities are implemented to help students learn? Is there a sound framework behind the structure of the course syllabus?</i>	<ul style="list-style-type: none"> • Teaching practices are not sufficiently planned or organized, or are poorly implemented • Practices are not well executed; little development in methods despite evidence of need • Students lack opportunities to practice the skills embedded in course goals • Student engagement is variable 	<ul style="list-style-type: none"> • Teaching practices are well planned and organized • Standard course practices carried out; follows conventions within discipline and institution • Students have some opportunities to practice skills embedded in course goals • Students consistently engaged 	<ul style="list-style-type: none"> • Activities are well planned, integrated, and reflect commitment to providing meaningful assignments and assessments • Uses effective, high-impact, or innovative methods to improve understanding • In- and out-of-class activities provide opportunities for practice and feedback on important skills and concepts • Students show high levels of engagement

<p>Achievement of learning outcomes <i>What impact do these courses have on learners? What evidence shows the level of student understanding?</i></p>	<ul style="list-style-type: none"> • Insufficient attention to student learning— quality of student learning is not described or analyzed with clear standards • Evidence of poor student learning; low level of skill/understanding is required or achieved without clear attempts to improve 	<ul style="list-style-type: none"> • Clear standards for evaluating the quality of student understanding • Typical student achievement for courses at these levels 	<ul style="list-style-type: none"> • Standards for evaluating student understanding are connected to program or curriculum expectations • Authentic assessments are used. • Quality of learning supports success in other contexts (e.g., subsequent courses or non-classroom venues), or is increasing within or across semesters
<p>Classroom culture and student perceptions <i>What are the students' views of their learning experiences? How has student feedback informed the faculty member's teaching?</i></p>	<ul style="list-style-type: none"> • Consistently negative student reports of teaching clarity, accessibility, interaction skills • Poor sense of learning among students • Little attempt to address concerns voiced by students 	<ul style="list-style-type: none"> • No consistently negative student ratings of teaching clarity, accessibility, interaction skills • Most students indicate progress with their learning • Instructor articulates some lessons learned through student feedback 	<ul style="list-style-type: none"> • Student feedback on teaching clarity, accessibility, interaction skills is generally positive • Evidence that classroom culture is respectful, cooperative, and encourages student motivation and engagement • Students perceive that they are learning important skills or knowledge • Instructor is responsive to student feedback in the short- and long-term
<p>Mentoring and advising <i>Has the faculty member worked individually with undergraduate or graduate students?</i></p>	<ul style="list-style-type: none"> • No indication of mentoring or advising students 	<ul style="list-style-type: none"> • Some evidence of effective advising and mentoring 	<ul style="list-style-type: none"> • Evidence of exceptional quality and time commitment to advising and mentoring

<p>Reflection and iterative growth <i>How has the faculty member's teaching changed over time? How has this been informed by evidence of student learning?</i></p>	<ul style="list-style-type: none"> • No indication of having reflected upon or learned from prior teaching or feedback 	<ul style="list-style-type: none"> • Continued competent teaching, possibly with minor reflection based on input from peers and/or students • Articulates some lessons learned from prior teaching and feedback 	<ul style="list-style-type: none"> • Regularly makes adjustments to teaching based on reflections about student learning (within or across semesters) • Examines student performance following adjustments • Reports improved student achievement of learning goals based on past course modifications
<p>Involvement in teaching service, scholarship, or community <i>In what ways has the instructor contributed to the broader teaching community, both on and off campus?</i></p>	<ul style="list-style-type: none"> • No interaction with broader community about teaching, including involvement with teaching-related committees • No evidence of keeping up with reports on effective teaching • Practices and results of teaching are not shared with others • Actions have negative impact on teaching culture in department or institution 	<ul style="list-style-type: none"> • Some involvement in teaching-related committees, or engagement with peers on teaching (e.g., teaching-related presentations or workshops) • Participates in department-level curriculum decisions 	<ul style="list-style-type: none"> • Regular involvement in teaching-related committees, engagement with peers on teaching (e.g., teaching-related presentations or workshops) • Occasional (or more) local or external presentations or publications to share practices or results of teaching • Contributes to department or university curricular planning or assessment • Advanced—Scholarly publications or grant applications related to teaching

Each dimension of teaching should have multiple sources of evidence in order to arrive at a consensus rating. These sources of evidence can be determined at the department level.

Possible sources of information for each dimension include:

- Goals, Content, and Alignment
 - Examples of exams and other assessments
 - Faculty self-report/portfolio/statement
 - Syllabi
- Teaching Practices
 - Faculty self-report/portfolio/statement

- Peer observations
- Student input (surveys/focus groups)
- Syllabi
- Achievement of Learning Outcomes
 - Examples of exams and other assessments
 - Faculty self-report/portfolio/statement
 - Student feedback related to learning goals
- Classroom Culture and Student Perceptions
 - Faculty self-report/portfolio/statement
 - Peer Observations
 - Student feedback related to learning goals
- Mentoring and Advising
 - Faculty self-report
 - Student input (surveys/focus groups)
- Reflection and Iterative Growth
 - Faculty self-report/portfolio/statement
 - Trends in numerical scores
- Involvement in Service, Scholarship, or Community Related to Teaching
 - External letters
 - Faculty self-report/portfolio/statement

Next Steps

As is clear from the description of the rubric and the underlying principles of having multiple lenses on the process of teaching evaluation, new measures that can be feasibly adopted by departments will be needed. There are ideas already emerging from colleges and departments on campus that can be adapted for use across campus as instruments for data collection, such as protocols for peer observations of courses. TEFD and OAPA can work cooperatively in consulting with departments and colleges as they work through using the rubric.

UMass is the lead institution on a 5-university grant from the National Science Foundation (NSF) on the subject of changing the evaluation of teaching and studying that change process (Gabriela Weaver, Co-Chair, TEFD, is Principle Investigator (PI)). The proposal, titled “Transforming the Evaluation of Teaching: A Study of Institutional Change,” was funded in late 2017. Three universities will engage in the work around changing the evaluation of teaching on their campuses (UMass, the University of Colorado at Boulder, and the University of Kansas), consisting of the adoption, adaptation, and sustainable use of new teaching evaluation approaches. A fourth PI (from Michigan State University) will study the process of transformation within and across the three campuses as case studies, focusing on what approaches work most effectively under what circumstances. An external evaluator (from Yale, via subcontract to UMass) will monitor the research aspects of the case study work.

For the UMass component of this project, the grant provides funding to support departments

that are interested in piloting the new rubric for teaching evaluation. It will provide two years of support for four departments piloting in the first two years of the grant (at \$5,000 and \$3,000 over the two years), and additional departments to receive \$3,000 in year three of the grant for additional piloting. After that, department adoption would be intended to proceed with less work required, since it will have the precedent of the pilot departments to build on, and therefore with no external funding.

REFERENCES

- AAAS. (2013). *Describing and Measuring Undergraduate STEM Teaching Practices: A Report from a National Meeting on the Measurement of Undergraduate STEM Teaching*. Washington, DC.
- Basow, S. A., & Martin, J. L. (2012) Bias in student evaluations. In M. E. Kite (Ed.), *Effective evaluation of teaching: A guide for faculty and administrators*. Society for the Teaching of Psychology.
- Benton, S. & Cashin, W. E. (2012). [Student Ratings of Teaching: A summary of research and literature](#). The IDEA Center.
- Bernstein, D. (2008). Peer Review and Evaluation of the Intellectual Work of Teaching. *Change : The Magazine of Higher Learning*, 40(2), 48–51. Retrieved from <http://doi.org/10.3200/CHNG.40.2.48-51>
- Chism, N. (2007). *Peer Review of Teaching: A Sourcebook* (2nd Ed.). Bolton, MA: Anker.
- Glassick, C. E., Huber, M. T., & Maeroff, G. I. (1997). *Scholarship Assessed: Evaluation of the Professoriate*. Special Report. San Francisco, CA: Jossey Bass Inc.
- Kinzie, J., & Kuh, G. (2017). Reframing Student Success in College: Advancing Know-What and Know-How. *Change: The Magazine of Higher Learning*, 49(3), 19–27.
- Kober, N. (2015). *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering*. Washington, DC: National Academies Press. <https://doi.org/10.17226/18687>
- Massachusetts Institute of Technology. *Educational Transformation Through Technology at MIT*. Retrieved January 3, 2017, from <http://web.mit.edu/edtech/casestudies/teal.html>
- National Academies of Sciences, Engineering, and M. (2017). *Undergraduate Research Experiences for STEM Students: Successes, Challenges, and Opportunities*. Washington, DC: National Academies Press. Retrieved from <https://doi.org/10.17226/24622.NASEM>
- Office of Academic Planning & Assessment, (2014). *Is there an “Innovative Instruction Effect” on Course Evaluation Results? Results from an exploratory study of Student Response to Instruction (SRTI) course evaluation results before and after Team-Based Learning Implementation*. Amherst, MA. Retrieved from (internal report)
- Office of Academic Planning & Assessment. (May 2016). *Internal report – analysis of SRTI scores with*

respect to faculty demographics.

- Spooren, P., Brockx, B., and Mortelmans, D. (2013). On the Validity of Student Evaluation of Teaching: The State of the Art. *Review of Educational Research* Season XXXX, Vol. XX, No. X, pp. 1–45 DOI: 10.3102/0034654313496870
- Stroebe, W. (2016). Why Good Teaching Evaluations May Reward Bad Teaching: On Grade Inflation and Other Unintended Consequences of Student Evaluations. *Perspectives on Psychological Science*, 11(6), 800–816.
- UMass Amherst. (2013). *Innovation and Impact: Renewing the Promise of the Public Research University*. Retrieved from <https://www.umass.edu/chancellor/sites/default/files/pdf/jtfsophase-i-report2.pdf>
- Uttl, B., White, C. A., & Gonzalez, D. W. (2016). Meta-analysis of faculty’s teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0191491X16300323>
- Wieman, C. (2015). A Better Way to Evaluate Undergraduate Teaching. *Change, January-Fe*(June 2012), 1–16. <http://doi.org/10.1080/00091383.2015.996077>

APPENDIX A: Known Methods for Evaluation of Teaching

A recent publication by the American Association for the Advancement of Science provides a general overview (not specific to science disciplines) of various methods that can be used for the evaluation of teaching (AAAS, 2013). These methods can serve as a menu of options for gathering feedback from multiple stakeholders with various viewpoints on the practice of a given instructor. The methods described in that document are summarized below.

Surveys

Surveys of students are by far the most common method in use nationwide to evaluate teaching. They are an efficient method for collecting evidence of the student perspective from a wide range of students across a large number of instructional contexts. However, the data they provide must be used appropriately given their limitations. The survey instrument must be appropriately designed and ask students questions that are both relevant to teaching effectiveness and focused on information students are capable of rating. In addition, the surveys must be appropriately administered with response rates and sampling methods that ensure the respondents are representative of the larger group. Because the quantitative data they provide can be easily scaled, standardized, and analyzed for trends across groups or over time, institutions may end up over-relying on these measures as the sole indicators of teaching effectiveness.

Surveys can also be given to faculty members themselves as part of their own teaching evaluation. Carefully designed surveys that have been described in the literature (AAAS, 2013; Wieman, 2015) allow faculty to report on classroom methods or course design elements that have been determined to be representative of high teaching quality. Surveys can also be used to probe faculty about learning goals for their course and to examine the connection between these and the overall outcome goals within a department or discipline.

There are a number of existing validated survey instruments, for either faculty or students, that probe beliefs, attitudes, and perceptions about teaching and learning generally (i.e., not specifically related to the course in question). These surveys can help establish baselines of understanding about the student expectations and faculty beliefs and practices.

Interviews

Interviews can explore teaching practices in greater depth, and can thus help to uncover causal relationships between practices and outcomes. Interviews can be done with either students or faculty, and can be carried out with individual subjects or focus groups. Because respondents can explain their thoughts, interviews may uncover new issues/ideas that would not emerge when responding to fixed questions on a survey. In this way, interviews also have the ability to more flexibly adapt to different situations in teaching. While interviews are commonly thought of as providing only qualitative data, it is possible to collect quantifiable data using interviews.

Interviews are not as simple to use for data collection because the interviewers must be well trained

and the data collection and analysis are time consuming. In addition, development of the interview protocol must be carried out in rigorous ways in order to ensure reliability and proper alignment of the data collected with the purpose of the instrument.

Observations

While surveys and interviews provide students' and instructors' own perceptions of teaching performance through self-reports, classroom observations offer the possibility of collecting direct measures of classroom teaching behaviors. Observations can help to document what approaches are used or whether or not particular approaches are being used in a classroom. As with interviewing, the observers must have specific training. Also, observations that are guided by specific protocols or rubrics grounded in the research on effective instructional behaviors can strengthen the validity, reliability, and standardization of the observational data. Observations that are structured by a protocol will focus the observer on looking for those aspects of teaching that are aligned with desired learning outcomes and with teaching effectiveness. Numerous examples of observational instruments exist in the literature, allowing departments to begin to adapt or wholesale adopt observations.

Videotaping can allow the observers to view the situation multiple times. Additionally, video data can be used in discussions with the observed faculty member to provide formative feedback. While observations have the benefit of providing an external lens on teaching evaluation, they are limited to documenting things that can be *seen* taking place in a classroom. Because many aspects of effective teaching happen outside of the class time (course design, syllabus, assignments, grading, faculty-student interactions), observations should accompany other measures that can provide insight into these aspects, such as interviews or document analysis.

Teaching Portfolios and Document Analysis

The materials that make up the course are often at the core of the knowledge development process for students. However, few teaching evaluation approaches provide a serious look at these materials and the role they play in a faculty member's pedagogy. Like observations, if portfolios are assembled with no guidance or structure, they will be inconsistent or lack clear connection to outcomes that the department or discipline deems important. In addition, such lack of guidance can result in portfolios that require unrealistic amounts of time to assemble, even as they provide little usable information. Nonetheless, portfolios can provide very rich insight into the content and functioning of a course and of the faculty member's composition of the overall experience. A good overview of how portfolios can be used is provided by Nancy Chism's book *Peer Review of Teaching: A Sourcebook* (2007) and an overview of literature on peer review of teaching can be found in Bernstein, 2008.

APPENDIX B: 2016 SRTI FORM

University of Massachusetts
Student Response to Instruction (SRTI)

PLEASE mark the bubble this way... ●
Use No. 2 pencil only.

This questionnaire gives you the opportunity to comment anonymously about this course and the way it was taught. Please reflect on your experiences in this course and answer each item below honestly and thoughtfully, filling in the circle for the response that most accurately reflects your experience. No results will be released until after grades have been filed.

Instructor's Name (please print)

1. The instructor was well prepared for class.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

2. The instructor explained course material clearly.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

3. The instructor cleared up points of confusion.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

4. The instructor used class time well.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

5. The instructor inspired interest in the subject matter of this course.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

6. The instructor showed a personal interest in helping students learn.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

7. I received useful feedback on my performance on tests, papers, etc.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

8. The methods of evaluating my work were fair.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

9. The instructor stimulated student participation in the class.

- 5 Almost always
- 4 Frequently
- 3 Sometimes
- 2 Rarely
- 1 Almost never

11. What is your overall rating of this instructor's teaching?

- 5 Almost always effective
- 4 Usually effective
- 3 Sometimes effective
- 2 Rarely effective
- 1 Almost never effective

12. What is your overall rating of this course?

- 5 One of the best
- 4 Better than average
- 3 About average
- 2 Worse than average
- 1 One of the worst

13. On average, how many hours per week outside of class did you spend working on this course (e.g., studying, reading, working on homework, group projects, etc.)?

- 7 More than 10 hours
- 6 Between 8 and 10 hours
- 5 Between 6 and 8 hours
- 4 Between 4 and 6 hours
- 3 Between 2 and 4 hours
- 2 Between 1 and 2 hours
- 1 One hour or less

14. The physical environment of the classroom was conducive to my learning.

- 4 Agree Strongly
- 3 Agree Somewhat
- 2 Disagree Somewhat
- 1 Disagree Strongly

10. Overall, how much do you feel you have learned in this course?

- 5 Much more than most courses
- 4 More than most courses
- 3 About the same as others
- 2 Less than most courses
- 1 Much less than most courses

15. 5 4 3 2 1 20. 5 4 3 2 1 25. 5 4 3 2 1

16. 5 4 3 2 1 21. 5 4 3 2 1 26. 5 4 3 2 1

17. 5 4 3 2 1 22. 5 4 3 2 1 27. 5 4 3 2 1

18. 5 4 3 2 1 23. 5 4 3 2 1 28. 5 4 3 2 1

19. 5 4 3 2 1 24. 5 4 3 2 1 29. 5 4 3 2 1

CLASS NUMBER

0	0	0	0	0
1	1	1	1	1
2	2	2	2	2
3	3	3	3	3
4	4	4	4	4
5	5	5	5	5
6	6	6	6	6
7	7	7	7	7
8	8	8	8	8
9	9	9	9	9

Which best describes this course for you?

- Requirement for your major/minor
- General Education requirement
- Other requirement
- Elective

What grade do you expect to receive in this class?

- A A-
- B+ B
- B- C+
- C C-
- D+ D
- F
- Other (e.g., taking the course Pass/Fail, auditing course)

What is your class level?

- Freshman
- Sophomore
- Junior
- Senior
- Graduate
- Other

Written comments about this course or your instructor are welcome. Please use the back of this form.

APPENDIX C: College of Natural Sciences Personnel Committee: Best Practices

College of Natural Science Personnel Committee: Best Practices University of Massachusetts Amherst

(Adopted as discussed at a CNS-PC meeting 3/31/2015)

This memo summarizes the best practices of the College of Natural Sciences Personnel Committee (CNS PC). We discuss expectations and procedures to be followed during consideration of reappointment, tenure, and promotion actions. It is important to emphasize that this is a living document, which will be subjected to continuous modification and reassessment as the CNS PC membership changes. Further, the Redbook is the official policy with regard to personnel issues (<https://www.umass.edu/provost/faculty-staff-resources/policies-contracts>). What is written here is not a substitute for the Redbook; rather, it is intended to provide an explanation of how the official policies laid out in the Redbook are implemented.

This memo is organized in the following 5 sections: (1) the timing of personnel processes, (2) committee operations; (3) the Redbook and specific operation of the CNS PC, (4) suggestions for preparation of documentation of cases by the Departmental PCs cases; and (5) additional considerations.

1. Timing of Personnel Processes

The CNS PC makes every effort to adhere to the personnel deadlines dictated by the Provost's Office. The Dean's Office provides the department chairs a timeline for preparation of cases and a checklist to ensure that personnel files are complete. In some cases, for example tenure, the timelines are so tight that preparation of cases should begin well before the semester in which the reviews are to be carried out (e.g., spring or summer for fall reviews, fall for spring reviews).

On occasion, a case file will be delayed. The CNS PC will make every effort to accommodate this delay. However, it is important for Departments to understand that it simply may not be possible to consider it. Late-arriving 4.2 and tenure cases cause particular trouble for the College, and command the (negative) attention of the Provost, Chancellor, and their staff. It is crucial that Departments adhere to the Provost's deadlines and deliver complete case files to the CNS PC on time.

2. Committee Operations

A. Committee Composition

It is expected that members of the CNS PC be tenured faculty members holding the rank of Full Professor. That being said, it is recognized that there may be circumstances under which a department decides that no full professors are available or that its most qualified representative is

a faculty member holding the rank of tenured Associate Professor. In such cases, it is requested that this department consult with the CNS PC, and explain the special circumstances. In the event

that an Associate Professor is on the CNS PC, this member may participate in the consideration of all cases as would any other member.

B. Cases involving potential Conflict of Interests (COI)

Departments will not select as a representative a faculty member whose own personnel action is to be considered during the term of appointment to the CNS PC, or shortly thereafter. Further, Departments will not select as a representative to the CNS PC a faculty member whose spouse or other close relation has a personnel action that will be considered by the CNS PC during the term of appointment, or shortly thereafter. The point of this guideline is to endure that no conflict of interest relationship can exist between any members of the CNS PC and a faculty member up for a personnel action.

In the event that there is the appearance of a COI of any kind by a given member participating in the evaluation of a particular case, those who might have any potential or perceived conflict of interest (co-author, co-PI, recommendation writer, present or past collaborator, etc.) will not be in the room for any part of the discussion of the case or for the vote.

Departments should also be mindful to not populate department PC committees with faculty who are in fact up for promotion for evaluation by DPC, and in general develop appropriate departmental COI safeguards.

C. Procedures

After a case folder has been delivered to the Dean's Office, electronic versions of the file are made available to PC committee members on a secure web site. Delivery of a complete file triggers the placement of the case on the agenda of the committee with a designated meeting date.

Primary and Secondary Reviewers are assigned to each case. The Primary Reviewer drafts a memo, with the collaboration of the Secondary Reviewer, discussing the strengths and weaknesses of the case. Prior to a full committee meeting, all other Committee members read the file, and the draft memo is distributed. At the full meeting, the Primary and Secondary Reviewers then lead the discussion of the case. The departmental representative member of the Committee is offered an opportunity to comment on the draft memo and to provide amplification or clarification as needed. The preferred role of the departmental representative is to provide clarifications and is not to serve as an advocate for the position taken by the Department – because the memos from the department should each make their case. The CNS PC urges that Departments prepare each file with adequate explanation such that it can stand alone on the merit content (see various suggestions for doing so, below). However, the departmental representative is invited to provide field-specific context, as needed.

The Committee will consider cases only if a quorum is present. A quorum is computed by taking half the total number of members, adding one, and rounding up to the nearest integer.

For tenure cases: Following discussion of the file, a straw vote is taken to assess the level of achievement (including votes of excellent, strong, or not strong) in each of the areas of research, teaching, and service and an overall vote on the award of tenure. If these votes reveal strong consensus, the vote is considered final – that is, unless an objection is raised by any member of the committee. The outcome of these votes is recorded, and the draft memo incorporates the vote outcomes and a summary of the issues raised during the case discussion. This letter is distributed to the entire committee for editing. The CNS PC chair then signs the edited letter and transmits it to the Dean’s Office. The committee considers such a case to be an uncomplicated case.

For promotion-to-full professor cases: Following discussion of the file, a straw vote is taken to assess the level of support for the promotion (there are not independent votes on the three traditional areas). As in the case for tenure, if these votes reveal strong consensus, the vote is considered final – that is, unless an objection is raised by any member of the committee. A draft memo then incorporates the vote outcome and a summary of the issues raised during the case discussion. As is the case for tenure, this letter is typically distributed to the entire committee for editing. The CNS PC chair then signs the edited letter and transmits it to the Dean’s Office.

For 4.2 cases: The discussion and subsequent memo should contain a careful assessment as to whether the candidate is on track for tenure in each of the areas of research, teaching and service. The memo should describe whether the candidate is on a trajectory to a positive tenure decision, and discuss what, if any, adjustments may be needed. The CNS PC memo for 4.2 reviews only includes the vote for reappointment. As with the tenure and promotion cases, the draft memo is distributed to the entire committee for editing. The CNS PC chair then signs the edited memo and transmits it to the Dean’s Office.

A complicated case results when the CNS PC reaches a consensus that does not agree with one or both of the Departmental recommendations. The Redbook specifies (section 6.4.f, page 15) that “Prior to a recommendation that may be contrary to either of the recommendations forwarded from the department level, the committee shall consult with the department.” This may involve, but is not limited to, an invitation to supplement their recommendations with additional information, and discussions with the Department Head and/or members of the Departmental PC who may be invited to a CNS PC meeting for this purpose. This consultation must take place before a final vote of the CNS PC is taken. Hence, this process may cause a delay in case resolution.

3. Comments on the Redbook and specific operation of the CNS PC

The CNS Personnel Committee takes the criteria for the various stages of promotion and tenure as enumerated in the Redbook at face value. Thus, we expect the Department to clearly document that each Redbook-enumerated criterion has been met. In tenure cases, where the Redbook calls for “excellence in at least two, and strength in the third, of the areas of teaching; of research, creative or professional activity; and of service,” we expect to find such excellence well- documented. Our understanding of individual cases is often related to the manner in which the case has been

prepared.

If relevant information is absent from the file, the committee will request that it be added before the discussion is concluded or a vote is taken. This may further delay action on a case. For example, it is important and helpful if a candidate's contribution to co-authored work and co-PI'ed grants be explained thoroughly to help the evaluation.

The CNS PC expects to receive documented evidence to support conclusions reached by Heads and Departmental Personnel Committees. Since CNS PC members are drawn from a wide variety of disciplines and change from year-to-year, they may well be unfamiliar with a candidate's discipline; thus, it is important that the written case analyses of the Departmental PC and the Head specifically discuss a discipline's criteria for evaluation and excellence. This is needed to help the CNS PC place a candidate's work in context. For example, in some fields conference publications carry great weight, while in others journal articles are the publication venue of record; in some fields, multiple collaborating authors are quite common, while in others they are less so; in some fields laboratory startups cause publications to appear later than in other fields; in some fields graduate student mentorship is the norm, while in others it is less common. The conventions regarding order of authors and senior authorship differ from field to field (and indeed sometimes within a field).

The CNS is a large college with diverse disciplines. While the Redbook is the official personnel policy document of the University, to attempt to understand the national norms for excellence and expectations among the various departments, the CNS PC has asked each department to provide to the committee a memo that outlines its "cultural standards." These so-called "cultural standards" are periodically updated. They serve to advise the committee members of the characteristics of excellence and expectations of faculty members, by indicating how the more general Red Book guidelines are interpreted by each of the departments.

To conclude this section we would like to comment specifically on two sections of the Redbook that are not often brought forward for discussion. Below are extracts from the two sections. The first of these seems implicit since promotion cases begin with Personnel Committees. The second seems to be not so universally recognized.

From Article III, section 3.1: "The faculty has primary responsibility in matters of faculty status, such as appointments, reappointments, promotions, tenure and salary adjustments."

And, from Article III, section 3.3: "No academic administrative official shall make a recommendation or decision which is counter to the original faculty recommendation without compelling reasons in written detail which shall specifically address the content of that recommendation as well as the established standards and criteria."

Please note the word "primary" in the quote from section 3.1 and the words "original" and "compelling" from section 3.3. Although undefined, it would appear that "original faculty recommendation" refers to the vote and memo from the Departmental Personnel Committee (DPC).

The quote from section 3.3 really means that unless the departmental recommendation from the DPC is negative the wording in the Redbook makes it difficult for a higher administrative level to overturn a positive recommendation from a department's faculty members (who, of course, know the discipline and discipline standards best). If a department is not happy with the record of a colleague based on its fair and honest evaluation of the body of evidence in the file, the department needs to utilize the appropriate professional standard and say "no" to the departmental colleague. Weak positive recommendations from departments, even if denied by the CNS-PC (and weak positive recommendations from departments are quite often denied by the CNS-PC, while citing its perception of compelling evidence) have traditionally not often been denied by administrative levels above the CNS-PC.

4. Suggestions for preparation of cases submitted for College PC consideration:

A) The File

A reappointment, tenure, or promotion case begins with the material provided by the candidate, typically consisting of a CV; personal/overview, research, teaching, and service statements; and selected research papers. In addition, the candidate should clearly describe in the personal/overview what the effort of the candidate has been with respect to Teaching; Research, and Service. For example, some positions require much larger efforts in Research, compared to Teaching or Service, while other positions may require a greater commitment to teaching.

Past experience has shown that it can be valuable for the department to mentor the candidate in preparing these materials, with one or more experienced faculty members (e.g., who have served on departmental and/or college PCs) reviewing and commenting on these materials. Some departments provide (with permission) to the candidate copies of statements from earlier cases. Of course, in the end it is the candidate who must determine the final set of materials submitted.

The CNS PC expects that each case folder submitted for consideration clearly presents the case for the faculty member under consideration. As noted above, the case folder should be in a form that can be properly evaluated by colleagues who are not familiar with the specific intellectual field of the department forwarding the case. To repeat, the case folder should include explicit documentation regarding the strengths of the candidate.

Departments help to ensure the optimal assessment of a case by showing clearly how the totality of the evidence leads them to their rating, but also by acknowledging those areas that are not as strong. It should be clearly stated what has the candidate accomplished such that the Departmental PC is convinced that the proposed action should be approved by the CNS PC. The committee recognizes that candidates who are appropriately judged 'excellent' by Heads and Departmental PC's may not show excellence in all aspects of each of the three traditional areas of their professional activities.

In all situations, context is important. So, for example, if a candidate has strong, but not excellent, numerical teaching scores, and these are offered without comment, the committee may conclude that the teaching case is strong, but not excellent. However, the actual teaching situation

may be such that the department deems such average scores as representing excellent teaching, perhaps due to the fact that the Department recognizes a class to be particularly challenging to teach. If the candidate's teaching scores are near a historic high for the relevant course(s), and the faculty member has received higher teaching scores for a course than just about anyone else, an explanation and some documentation of this would help the CNS PC understand and appreciate a rating of excellent by the department. Similarly, a course may have been particularly difficult to teach because it was new, or the instructor was assigned to the course at the last minute. Along these lines, it helps the committee if the department provides a description of the normal load and variety of courses expected of faculty in that department for the particular promotion under consideration. Teaching loads vary, and some departments expect faculty, including junior faculty, to teach well at all levels (non-major, major & graduate) and to both small and large classes, while other departments do not.

Similarly, departments should comment on normal service contributions and expectations at the given stage, noting expected levels of service to the department, college and/or profession; as with the other areas, the achievement of strength or excellence should be well-documented.

Finally, the CNS-PC and the Dean's Office urge that in the memos from the Department Personnel Committee and from the Head/Chair no names of referees be used. Thus, instead of saying that "Professor X (location)" said this or that, it is far preferable to simply say, "one referee" or "a referee" with no other identifier. Also, there is no need to further identify a referee as "on campus" or "off campus" or "close" or "not close" in the text of the recommendation memos.

B) Letters from Reviewers:

Tenure and Promotion-to-Full : External letters that document the stature and impact of the faculty member and his or her research program are a crucial and required component of tenure and promotion-to-full cases. Consequently, it is suggested that a significant number of the external letters come from those with Full Professor or equivalent seniority and status. By tradition, typically, half of the supplied letters come from individuals suggested by the candidate, and half as determined by the department. Of the letters needed for tenure and promotion, no more than half should be from official reviewers close to the candidate.

Effective with considerations to be made in the 2015-2016 academic year, in most cases dossiers should include 10-12 external letters from leaders with significant stature in the discipline. This Provost Office guideline has moved closer to our own current CNS guidelines. The CNS-PC recognizes that in some disciplines it is simply not possible to find this number of highly qualified referees. In cases where it is not possible to obtain this many letters, a clear explanation of reasons why and the efforts made should be presented. The goal should be that the candidate file not contain less than eight external letters.

Reviewers close to the candidate are considered to be: graduate supervisors or mentors; dissertation committee members; co-authors, co-PIs, or other collaborators; former fellow graduate students, etc., but not necessarily individuals met at a conference, those that the candidate sat with on a national committee/panel, or a fellow journal editor, for example. Letters

may be solicited from collaborators, as well, who have worked closely with the candidate and frequently either published or garnered research funding together. These letter writers should be asked to comment on the nature of the candidate's contribution to the collaboration. Letters such as these may be identified as those of official reviewers, with a clear designation that they are from persons who are "close" to the candidate.

Letters from internal (UMass) faculty and research staff are also common, but do not replace the need for external letters. They are particularly valuable in cases where the candidate has worked closely with the letter writer on joint research, teaching, or service activities. Typically the expected number of such letters is no more than five unless there are unusual circumstances – e.g. many on-campus collaborations.

The personnel file should include a list of the names, addresses, and a brief description of each reviewer who was asked for a letter, including whether they are "close" to or "not close" to the candidate. The brief description of each reviewer should be detailed enough (e.g. a few sentences) to establish the credentials, stature and relevance of the referee. A copy of the formal letter requesting a review should be included in the personnel file. We expect that such solicitations be explicit about the nature of the personnel action. The CNS PC requests that Departments do *not* include information about referees who have not replied or who have declined to write a letter.

We wish to emphasize that, as noted in the Redbook, sections 4.1, 4.6 and 4.9, there are really three key provisions that are relevant to the grant of tenure: excellence in two of the three traditional areas, recognition on and off campus among scholars in the discipline, and clear evidence for continuing professional development and achievement. According to the Redbook, all three conditions must be met for the tenure recommendation to be positive.

4.2 Review: Not all departments in the current CNS have a past practice of asking for external letters as part of the 4.2 review. Further, we recognize that in some disciplines it has previously not been the practice to ask for external review at the pre-tenure stage. Because of this we did not require that a 4.2 case include external letters, although such letters were always welcome because they informed the department and the CNS PC and often provided very valuable guidance to the faculty member under review. For departments that routinely asked for letters at the 4.2 stage, we recommended that they obtain three to five external letters. As with tenure and promotion cases: (i) the letters are expected to include reviewers who are selected by the Department (rather than the candidate) and reviewers who are not close to the candidate, and (ii) the file should include the same explanatory information about the reviewers as for tenure and promotion cases. Again, internal letters from UMass faculty may be valuable and are welcome.

After discussion with departmental representatives the CNS-PC has concluded that such letters provide such valuable guidance to candidates that they should be required. Consequently, we urge that departments that previously did not solicit external letters do so in the future for the following positive functions of such letters. Letters obtained for 4.2 reappointments function as a mechanism to help tenure track faculty establish an even better record for tenure review. Such letters may be from mentors, colleagues and academic friends as well as independent outsiders. It should be recognized that letters from reviewers "not close" to the candidate may provide very useful insights and perspectives that are not necessarily apparent from inside the department.

Solicitation of these letters increases the visibility of the candidate's work. It may also occur that there is no internal expert in the candidate's subfield of research. In this case, it is not possible to conduct a purely internal review that accurately evaluates the candidate's research and progress towards a tenure-quality research program. External letters both protect the candidate and add important information.

The CNS PC stresses the value of *the feedback provided to the faculty candidate* that is provided by such external letters in the 4.2 process. The 4.2 process is a mid-course review on the way to tenure and while the review may result in non-reappointment, the great majority of such reviews result in helpful feedback and advice to the candidate. This advice is substantially enhanced by the presence of such letters. Thus, the CNS PC very strongly encourages that *all departments* include such external letters as a natural part of the 4.2 review.

The committee recognizes that early in a new faculty member's career it may be difficult to evaluate achievement due to a limited opportunity for publication caused by the initiation of a research program. But, visibility at national meetings, for example, may allow the growing stature of the faculty member to be recognized. Additional evidence for a well-launched research program includes invited presentations at national or international meetings based on work carried out since the faculty member was hired, publications, and other forms of recognition based on recent performance. The receipt of external peer-reviewed grant awards enhances the belief that a faculty member will have the ability to carry forward his or her research. Similar evaluation of teaching and service accomplishments are required. A major goal of the 4.2 review is to evaluate the progress, offer constructive feedback on that progress, and provide an early experience with the performance review process.

5. Additional Considerations

Teaching evaluation: An excellent teacher can be recognized from a variety of evidence. The candidate's teaching statement and submitted materials/portfolio are carefully considered. At UMass, SRTI scores with comparative departmental and university averages are currently the standard numerical evidence of teaching quality (http://www.umass.edu/oapa/srti/pdf/interpreting_srti_results.pdf). It is important to assess response yields, dispersion in the numerical scores, and comparative scores generated for a given course taught by the candidate and by other faculty. If there are insufficient comparisons for the same course, comparisons with similar courses can be shown (and the sense in which those courses are 'similar' should be discussed). The CNS-PC recognizes that such scores are not perfect, their value can be debated, and they alone do not give an adequate picture of a faculty member's teaching quality. However, they do allow the committee to see the evolution of student evaluation over time for repeated courses and they are one measure that is for the most part used University-wide.

Many departments have adopted a mode of presentation of SRTI scores that is very helpful to the CNS PC and we strongly encourage all departments to summarize the SRTI scores from courses taught and present them in tabular form. So, *in addition* to the submission of the standard SRTI

results with scores and comparison histograms, we encourage submission of the SRTI scores in tabular form along the following lines.

SRTI Q#	Semester #		Semester #		Semester #	Etc.	
	Course A	Mean	Course B	Mean	Course C		
response							
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							

In the table, columns should record the SRTI scores for each course and in the same column (or a separate one) the standard deviation of the score. In the table “Mean” refers to the mean score for the collection of similar course-level courses taught by the department that semester. In the “response” row, a measure of the response rate should be reported, e.g. 82%, for each course evaluated. It is also helpful for the table in that row to include the number of registered students in each course.

Beyond these ratings, various other evidence of teaching strength such as signed letters from students have taken classes with the candidate, letters from undergraduate or graduate research advisees, letters from colleagues who have co-taught courses or sat in on classes, evidence for participation in innovative teaching programs, and documentation related to major initiatives in course re-design or pedagogical advances should be included. Letters from students, in particular, help with assessment of the SRTI scores. Also helpful to the CNS PC are comments about how the SRTI scores for the faculty member under review compare with the mean scores obtained by previous instructors who have taught the same course. As noted above, personal statements from the candidate that discuss aspects of teaching excellence are essential.

The CNS-PC expects that in the case of student letters, these will be solicited from the span of student-faculty interactions that are relevant. So, for example, to primarily submit only letters from students who have been mentored by a faculty member in a laboratory setting would provide an incomplete view. The CNS-PC expects to see individual letters solicited from students who have experienced lecture-style classes with the faculty member. Many departments solicit such letters from a random distribution of students in large classes and all of them in small classes.

Access to such letters provides the CNS-PC with a much better ability to evaluate teaching than would the SRTI scores alone.

Interdisciplinary and collaborative research. The committee is aware of the importance of collaborative, interdisciplinary approaches to many research questions. In no way should the need

for evaluation of the candidate discourage involvement in such research endeavors. Nonetheless, the necessity to evaluate the candidate's research stature and accomplishments requires that his or her contribution to partnerships must be made clear. Thus, the department should provide adequate explanation and documentation to enable the committee to evaluate the candidate's accomplishments based on collaborative research work and to describe clearly how the department evaluated such work.

Authorship. The committee recommends that the department explain the usual practices in the candidate's field regarding order of authors. Additionally, the department and/or candidate should identify co-authors who are students or post-doctoral fellows working directly under the candidate.

Confidentiality. The CNS PC maintains confidentiality during the evaluation process. We do not discuss details of the case with those outside of the committee. We do expect, however, that departmental representatives will provide appropriate feedback to Heads and Departmental PC Chairs, information that will better allow them to represent their faculty members. For example, in cases where the committee has requested that a departmental chair or PC chair attend the CNS PC committee meeting to discuss a case, it would be appropriate for the department representative from the CNS PC to indicate to the Head and PC Chair the context that motivated the invitation, i.e., to indicate areas in the case where questions had arisen. The CNS PC Chair may also help to provide the context.

Early cases. As outlined in the "Red Book," Assistant Professors are routinely considered for tenure in their sixth year of service. Similarly, there is a generally accepted, typical additional time for promotion to Professor (usually 5 or 6 years); this is not a rule, simply past practice and tradition. The CNS PC carefully scrutinizes cases that are early to ensure that adequate evidence is in place to support the pending action. However, the CNS PC also believes that the case is ready when the case is ready, and does not view an early case with prejudice. There is ample precedent for early personnel actions and in early cases approved by previous committees the evidence has been compelling and well documented.

New hires. In cases in which a new faculty member is hired with the recommendation that tenure be granted, the committee takes the position that the case is to be evaluated in the same manner as would be the case for a faculty member who is already here. Thus, the practice of the committee is to apply the same standards, and to expect the same level of documentation that supports assertions of excellence as would be the case for a faculty member who has been at the University for six years. In particular, for newly hired faculty members, the CNS PC expects to find a documented case in the areas of teaching, research, and service.

Late cases. There are situations in which faculty members who have been on the campus for many years are put forward for promotion well beyond the standard interval since the last promotion, sometimes near the end of their careers. We approach each case with the same philosophy: the case is ready when the case is ready. But, we also recognize that when it is not, it is not. So, a late case that does not contain the appropriate evidence that would be considered adequate for a more timely promotion will likely be denied by the CNS PC since the committee

attempts to maintain appropriate standards consistent with the Redbook in all cases, regardless of years in rank.

Tenure (or promotion) Cases in situations in which a more senior faculty member is hired without the initial grant of tenure. There are occasional situations in which a more senior investigator is hired (e.g. from Industry or a national lab) at an advanced rank but without tenure. This can happen, for one example, when there is no prior teaching experience. Eventually tenure (and/or perhaps promotion) will be considered and the question emerges as to how much emphasis should be placed on productivity while at UMass relative to previous productivity.

Conversation with the Dean (in the spring of 2014) resulted in the following view being adopted by the committee. In cases of extensive research experience prior to arrival at UMass, that prior work can be evidence of meeting the criterion of establishing independence in the research program and can contribute to the overall assessment of research quality. While such prior work can continue to be relevant, the CNS PC and the institution also wants to see clear evidence that the promise presented when the person was hired has met expectations - and has led to continued productivity. Productivity while at UMass is important because it is the key basis for confidence that there will be a growing body of work in the future. In terms of the Academic Personnel Policy (Red Book), work at UMass must give “reasonable assurance of continuing development and achievement leading to further contributions to the University.” In the end, we are to decide if a convincing case for excellence has been made in the areas in question and thus if the portfolio is adequate for the award of tenure.

The committee view is that, while prior work can be relevant—particularly for evidence of an independent research program, its relevance for assurance of continuing development diminishes as one goes further back into the past. Consequently, our interpretation of is that the most recent probationary period (at UMass) is most important for assurance of continuing development, but some consideration can be brought to bear from prior work.

Closing Comment: To close, we emphasize that this memo is offered in good faith as a mechanism to clarify practices of the CNS PC, and to focus our collective attention on items that assist the committee as it does its work. The interpretation of documentation, evidence, and even criteria for often-used words such as "excellence" are in the end up to the individual faculty members on the CNS PC. These faculty members are selected for, among other things, their experience, fairness, and judgment. The membership of the committee, like that of most Departmental Personnel Committees, changes somewhat from year to year. Thus, it is only reasonable to expect that there will be modest changes in the manner in which the individual members of the CNS PC make their decisions. Nevertheless, we endeavor to maintain a consistency in our procedures that will allow all concerned to prepare personnel actions effectively and consistently. Personnel actions have long-term impact on both the future of the individual candidates, and the future of the University. Hence we have a responsibility to ensure that decisions are made with care.



INSTRUCTOR'S FORM

PEER-to-PEER Teaching Review

Peer-to-peer teaching reviews are intended to offer an opportunity for constructive conversation about teaching. Below is a tool intended to help document an observed class. You should adapt the questions or areas of consideration to fit your circumstances. Ideally, peer observation allows an instructor to identify the kind of skills, practices and procedures that can be improved over time. For that reason, it is important to think about the class in terms of the instructor's intent. For the purpose of this review, the instructor will provide information about their course in the Instructor's Form and the Reviewer will record their observations on the Reviewer's Form. These two forms should become the basis of a conversation after the observed class.

I. PEER INFORMATION

Name of instructor _____ Department _____

Title & Course # of class being observed _____

Name of observer _____ Department _____

Number of Students _____ Date and Time _____

II. REFLECTION ON TEACHING

1) What are your strengths as a teacher? ¹

2) How would you like to improve as a teacher?

¹ Adapted from the Academy of Teaching and Learning Fellows, University of Cincinnati.

PEER-to-PEER Teaching Review

Peer-to-peer teaching reviews are intended to offer an opportunity for constructive conversation about teaching. Below is a tool intended to help document an observed class. You should adapt the questions or areas of consideration to fit your circumstances. Ideally, peer observation allows an instructor to identify the kind of skills, practices and procedures that can be improved over time. For that reason, it is important to think about the class in terms of the instructor's intent. For the purpose of this review, the instructor will provide information about their course in the Instructor's Form and the Reviewer will record their observations on the Reviewer's Form. These two forms should become the basis of a conversation after the observed class.

I. PEER INFORMATION

Name of instructor _____ Department _____

Title & Course # of class being observed _____

Name of observer _____ Department _____

Number of Students _____ Date and Time _____

II. CLASS OBSERVATION

1) Describe the presentation of information by the instructor. What methods did the instructor draw upon to begin the conversation? Did the instructor assume student reading knowledge or written exercises?¹

2) Describe the student-initiated discourse.

¹Adapted from the Academy of Teaching and Learning Fellows, University of Cincinnati.

