

Assessing Teaching and Learning at W&M
Academic Affairs Committee of Faculty Assembly
January 6, 2021

EXECUTIVE SUMMARY

During AY 2019-20, the Academic Affairs Committee of Faculty Assembly conducted a review of how we evaluate teaching at W&M using four approaches: literature review, a survey of current W&M practices, a review of current W&M policies, and case studies of recent reviews at other institutions. We chose to focus on this particular issue primarily because: (1) it is a topic that is frequently raised as a concern in Faculty Assembly surveys of the W&M faculty (e.g., 2016, 2020), (2) it is a topic of recent discussion in all-school faculty meetings (e.g., A&S) as the COLL curriculum introduces an equity and social justice requirement, and (3) it is one component of a larger discussion on how faculty at W&M are evaluated in general, through the merit evaluation process.

This document summarizes our methods, findings, and recommendations, with raw data and information provided as appendices.

Given the extreme variation in teaching definitions, expectations, and modalities across W&M schools, there is no single approach to evaluating teaching that will be applicable across all of the units. Instead, we recommend that each school develop a committee (to include administrators, faculty, and students) to assess these recommendations and to determine how they can be implemented more effectively within the culture of that school. Ideally, we recommend that implementation take place AFTER COVID-related procedures have either been lifted or normalized.

Our proposed timeline is as follows:

- **By Summer 2021:** Provost works with Deans from each of the schools to delineate which aspects of the process will be determined at the university versus school- level.
- **Fall 2021:** Each school initiates its own implementation committee to include faculty, students, and administrators
- **By Summer 2022:** Each school has: (1) developed a definition of effective teaching and teaching expectations, (2) established a clear statement of importance of teaching relative to competing priorities, (3) communicated this definition and statement school-wide, (4) revised policies and procedures related to assessing teaching, and (5) updating relevant handbook policies.
- **Starting Fall 2022:** Each school begins pilot programs and implementation.

Our major recommendations are outlined below:

- Elements of the teaching evaluation process should be designed centrally and adopted consistently across units but the process must be customized within schools as needed so that it can reflect the unique missions and goals of each school, as well as their unique student populations.
- Develop a working, and comprehensive, definition of effective teaching and teaching expectations for each school

- Develop a clear statement of the importance of teaching relative to competing priorities, i.e., research and service for each school.
- Communicate this definition and statement to the school and university communities, including students, faculty, and administrators. Make sure that guidelines are clearly communicated as a function of career stage (e.g., pre-tenure), tenure track status, maintenance of “research active” status and what that entails, etc.
- Increase the focus on *formative*, as opposed to *summative* evaluation. Summative evaluation seeks to produce judgments about whether an educator is effective or not (often for merit review), whereas formative evaluation focuses on *improving* teaching. Evaluation should be viewed as a *developmental* practice that goes beyond simply evaluation and explicitly helps educators improve their effectiveness.
- Shift the terminology used across the schools from *evaluation* to *assessment*, to signal a more formative, and developmental, approach to this process.
- Encourage a growth mindset (among students, faculty, and administrators) that considers evaluation a positive experience linked to the possibility of improvement and growth. Traditional review processes that emphasize evaluation over development are often interpreted by educators as punitive or preventive. One consequence of this approach is that it can instill a fixed mindset with regard to teaching, one in which teaching ability is permanent and beyond improvement.
- In addition to student feedback, the annual evaluation of teaching at W&M should include approaches BEYOND student course evaluations (which, as almost all studies demonstrate, disadvantage faculty based on race/ethnicity, gender, sexual orientation, career stage, and more). Approaches could include peer observations (e.g., evaluators within units, across units, or even external consultants), feedback from administrators or program leaders who can assess contributions outside the classroom (e.g., curriculum development, advising), self-assessment (e.g., a narrative about teaching philosophy, contributions, assessment of progress towards past goals, future goals, etc.), teaching portfolios, peer or administrative assessment of course materials, peer or administrative assessment of student work, and engagement in workshops, projects, and academic innovation efforts. Assessment should involve *triangulating* sources of information (finding multiple sources at a point in time that corroborate or qualify a finding) and looking for patterns and change *over time*.
- Eliminate unidimensional, “global” student course evaluation ratings – especially with respect to the individual educator (vs. course). Instead, develop questions based on established, multi-dimensional conceptualizations of teaching effectiveness, including the following: Learning/value, Enthusiasm, Organization, Group interaction, Individual rapport, Breadth, Examinations/grading, Assignments, and Difficulty.
- Use of student course evaluation data should avoid the use of mean scores without measures of distribution (standard deviation, range, etc.) and explicitly address issues of:
 - reliability (e.g., sufficient sample size and random sampling)
 - validity (are we measuring what we think we are)
 - fairness (what are we comparing the scores to?)
 - consequences (how might educators respond to ratings? might they “game” the process in a way that inflates ratings but hurts the learning experience?).
- Make distinctions between programmatic differences, differences in student type or stage, disciplinary expectations for learning, and how these relate to the effectiveness of instruction. Schools could release anonymous data on course evaluation scores for similar

types of courses across units (e.g., COLL courses, capstone courses, large lecture courses, etc.) to give faculty an opportunity to assess their teaching in a broader context.

- Steps should be taken to understand and assess the ways in which faculty engage in teaching outside of the classroom (including curriculum development, advising, research mentoring, and exploration of innovative teaching approaches), including quantifying the commitment that this kind of engagement entails.
- The process should, by definition, be developed by those who will be involved (i.e., including faculty and students).
- To avoid overwhelming the Procedural Review Committee with the school and unit-level changes to Faculty Handbook text that will result from this process, we also recommend that the Provost works with the Deans of each school to streamline the process and ensure that implementation is not delayed.

We view this analysis as a first step in strategically reassessing how W&M defines and assesses excellence in teaching.

TABLE OF CONTENTS

Abbreviations.....	3
Committee Membership.....	3
Methods.....	3
Literature Review.....	5
Survey of Practices.....	12
Review of Policies.....	17
Case Studies.....	26
Appendix A: Survey of Practices-- Survey Instrument.....	29
Appendix B: Survey of Practices-- Quantitative Data Summary.....	42
Appendix C: Survey of Practices-- Qualitative Raw Data.....	46
Appendix D: Case Studies-- Interview Questions.....	63
Appendix E: Case Studies-- Raw Data.....	64

ABBREVIATIONS

A&S: Arts & Sciences; NTE: non-tenure eligible faculty; RPT: Retention/Tenure/Promotion; SET: Student Evaluations of Teaching; SMS: School of Marine Sciences, Virginia Institute of Marine Science; SOB: Raymond A. Mason School of Business; SOE: School of Education; SOL: Marshall–Wythe School of Law; STLI: Studio for Teaching & Learning Innovation; TE: tenure-eligible faculty.

COMMITTEE MEMBERSHIP (2019-20)

Rowan Lockwood (Chair, A&S), Mark Brush (SMS), Lynda Butler (SOL), Samantha Easby (doctoral student, SOE), M. Brennan Harris (A&S), Natoya Haskins (SOE), Mark Hofer (STLI, SOE), William Hutton (A&S), Michael Luchs (SOB), John Riofrio (A&S), Brad Weiss (A&S)

METHODS

Our approach to exploring how we evaluate teaching at W&M involved four main components as outlined below:

1. **Literature Review:** We conducted a review of the published peer-reviewed literature on most effective practices in evaluating teaching in higher education. To do this, we focused primarily on review articles published in the past decade that summarized the major trends, challenges, and recommendations. We also compiled additional references from interviews and materials developed by other institutions investigating similar questions (Case Study approach, see #4). Finally, we reached out to W&M contacts (Mark Hofer, Pamela Eddy, and James Barber) from the SOE for their recommendations. The literature review was performed in Fall 2019 by Michael Luchs and John Riofr
2. **Survey of Practices:** In order to determine how W&M evaluates teaching in practice, we developed a Qualtrics survey of 26 questions (Appendix A), focusing on the specifics of how units evaluate teaching. Questions delved into what information is used to evaluate teaching (i.e., SETs, peer classroom observation, self-reflection, etc.); the extent to which activities outside the classroom are included (e.g., course development, teaching innovation, individual student support, etc.); faculty views on the effectiveness of current approaches for faculty evaluation, faculty development, curriculum development; the emphasis and application of SET scores and comments, and more. The survey was developed by the committee (with substantial input from Rowan Lockwood, Natoya Haskins, and Michael Luchs), piloted with 4-5 units, and opened from February 7-March 10, 2020. The survey was completed by the Deans (or Associate Deans) of the SOE, SOB, and SOL, because the evaluation of teaching in these schools does not vary significantly across units within these schools. For SMS and A&S, the survey was sent to ALL chairs and program directors because units at these two schools take different approaches to evaluating teaching. This approach yielded detailed data for each of the five schools, as well as all four SMS units, and a representative sample of 19 A&S units including: Anthropology; Art; Art History; Asian and Middle Eastern Studies; Biology; Classical Studies; Economics; English; History; Linguistics; Mathematics; Modern Languages and Literature; Music; Neuroscience; Philosophy; Physics; Psychological Sciences; Sociology; and Theatre. The A&S units represent all three content areas, both departments and programs, and both undergraduate only and graduate degree-granting programs. Quantitative compilation of the survey results was performed by Brennan Harris. Qualitative analysis of the survey results was performed by Natoya Haskins.
3. **Review of Policies:** To explore the extent to which W&M policies dictate our approach to evaluating teaching, we performed a review of policies across all five schools, including several units within A&S in the Spring of 2020. We obtained copies of policies, including any information related to annual merit evaluation, retention/tenure/promotion, and faculty development from the Provost's office and from the Deans' offices for each of the five schools. We were able to compile unit-level policies for A&S from 23 different units, representing all three content areas, both departments and programs, and both undergraduate only and graduate degree-granting programs. These units were: American Studies; Anthropology; Applied Science; Art and Art History; Biology; Chemistry; Classical Studies; Economics; English; Government; Health Sciences; History; Linguistics; Mathematics;

Modern Languages and Literature; Music; NIADH; Philosophy; Physics; Psychological Sciences; Public Policy; Religious Studies; and Theatre, Speech, Dance. Review of policies was performed by Bill Hutton, Mark Brush, and Lynda Butler.

4. **Case Studies:** To determine how other higher education institutions evaluate teaching, we conducted a case study approach, by reaching out to several universities that have recently embarked on a reassessment of how they evaluate teaching. We reached out to eight universities, representing public and private, small and large, across different geographic regions. The universities we were able to interview were: Smith College, University of Kansas, University of Oregon, University of Southern California, and Vanderbilt University. We requested any information on their review of teaching evaluation that they were willing to share, and then conducted in-depth phone or Zoom interviews in February 2020. The list of interview questions is provided in Appendix D. Case study work was performed by Brad Weiss, Samantha Easby, and Rowan Lockwood.

LITERATURE REVIEW

This report provides an overview of key issues with regard to the *evaluation & development of teaching & educators* at William & Mary. In particular, this report will highlight the various approaches to evaluating post-secondary teaching (both at William & Mary and across the U.S. generally), identify significant problems with the most common approaches, foreground the issues relevant to the management and execution of teaching evaluations, and offer some insight into alternative approaches.

Survey of Literature

The efficacy of teaching evaluations has been a topic of significant, and sustained, research for at least the past four decades. This ample body of research offers insight into the complexity of measuring teaching effectiveness, analysis of the various strategies developed for quantifying teaching effectiveness, as well as diagnosis of the flaws associated with current and ongoing practices of teaching evaluation and development. We organize our synthesis into three general categories: **Purpose & Philosophy, Approach, and Management & Execution**. While there is some conceptual overlap between these sections, we believe it is useful to address each in turn given the complexity of the topic (e.g., it is easier to consider a future approach when one is grounded in the Purpose and Philosophical perspective).

First, some brief terminology:

- Student Evaluations of Teaching, or SET, is the same as
- Student Rating of Instruction, or SRI

PURPOSE AND PHILOSOPHY

Defining “effective teaching:”

The problem: Many institutions, including W&M, place significant emphasis on the importance of effective teaching. However, extensive research into teaching evaluations reveals unstated assumptions about what constitutes effective teaching. Often no distinctions are made between programmatic differences, differences in student type or stage, or differentiation

between disciplinary expectations for learning and how these relate to the effectiveness of instruction.

Recommendations: At W&M, enormous weight and significance is given to SET teaching scores, in particular, the question (or some derivation) “rate the overall effectiveness of the instructor.” Because there is no clear university-wide definition of what constitutes “effective teaching” for either students, faculty, or administrators, it is important that the university consider ways to convene interested constituents in order to come to a working, and comprehensive, definition of effective teaching. Furthermore, efforts should be made to inform the university, including students and administrators, of the working definition.

Purpose of the Review process

The problem: Institutions, by and large, have implemented review practices that focus exclusively on the summative evaluation of teaching in order to make material decisions on promotion and compensation. However, current research emphasizes the importance of evaluation as a *developmental* practice, one that goes beyond simple evaluation and, instead, has the stated goal of helping educators improve their effectiveness.

Recommendations: Any revision to the teaching evaluation process should consider the difference between *summative* and *formative* evaluation. Because summative evaluation seeks to simply render judgments about whether an educator is effective or not, and because the evaluative process is often hampered by non-objective measures, it is important to consider multiple modes of evaluation. Emphasis should instead be placed on formative evaluation whose purpose is focused on *improving* teaching. This means that from the outset the process is understood as an ongoing opportunity for refining methods, strategies, and outcomes (Benton & Young, 2018).

Scope of Review Process

The problem: Institutional practices of teaching evaluation often default to excessively narrow definitions of what constitutes “teaching.” These tend to focus exclusively on traditional classroom practices. This approach effectively ignores subtle and important teaching moments that occur outside the confines of the classroom and that include, among other things, curriculum development, advising, and mentoring (both formal and informal). In the case of W&M faculty, this includes teaching/mentoring that takes place in research labs, in theater/dance rehearsals, individual advising, office hours, and advising of student research of affinity groups. In particular, because of a lack of diversity among faculty, another source of hidden workload is the unofficial mentoring of students of color that frequently falls to faculty of color (Fryberg & Martínez, 2014).

Recommendations: To address this, steps should be taken to understand the many different ways and contexts in which faculty engage in teaching outside their classes. Secondly, efforts should be made to quantify the time and commitment that this kind of engagement entails. Thirdly, teaching evaluation measures should include a recognition and assessment of these alternate forms of teaching. This might include self-assessment by faculty as well as brief statements of impact from students.

Focus on the person rather than behaviors and behavioral change

The problem: Educational institutions frequently employ evaluative processes that implicitly conceptualize effective (or ineffective) teaching as a set of fixed traits. In this scheme, one is

either a good instructor or a bad instructor. When the review process is communicated in this manner, it hinders the possibility for educators and administrators to understand teaching as a set of practices that can be adapted, refined, and improved upon.

Recommendations: The most current research on teaching evaluations makes an urgent appeal for broadening the kinds of materials used in order to evaluate an educator's performance. In addition to SET scores (which continue to be seen as fraught with potential bias), experts recommend including peer evaluations, self-assessment, and portfolios (ASA, 2019; Berk, 2018; Benton & Young, 2018).

The Evaluative Culture: Fixed versus growth mindset

The problem: Traditional review processes that emphasize evaluation over development are often interpreted by educators as exclusively punitive or preventive (Berk, 2016). Because evaluations are often narrow in scope (focusing exclusively on SET) and also linked to high stakes such as review for promotion and merit review, faculty are often reticent to see the *developmental* value in evaluations. This contributes to a sense that undergoing an evaluation of teaching becomes fraught, antagonistic, and frequently demoralizing. Under these conditions, educators and administrators can come to an unacknowledged understanding of teaching as a skill that educators either do well or do badly. One important consequence of this framework is that it has a tendency to instill in educators the notion that teaching ability is fixed, or permanent, and consequently beyond improvement and in which innovation is seen as an unnecessary risk to positive teaching evaluations.

Recommendations: Floden (2017), in an extensive meta-study of educators' attitudes towards evaluations, observes that educators are more likely to see benefit in the evaluation process when they feel that it will reveal issues, within their control, that they can work on and improve. Floden (2017) also emphasizes the role of creating a culture (among students, faculty, and administrators) that considers evaluation a positive experience linked to the possibility of improvement and growth. Efforts to improve the evaluative process should strongly consider how to (re)educate the university community on the purpose and potential for evaluation.

APPROACH

Educator (and teaching) expectations

The problem: Many institutions suffer from vague, inconsistent, or incomplete expectations about the role of teaching (i.e., how important it is to the institution), teaching workload (as a function of career stage, tenure track status, research productivity, etc.), how teaching will be evaluated and rewarded, and even about what constitutes "teaching" (e.g., just course delivery, or also course development, program development, student advising and mentoring, etc.).

Recommendations: Develop and communicate a clear understanding of expectations - likely specific to each School or Area within W&M - about the role and importance of teaching with respect to the competing priorities of research and service. Make sure that guidelines are clearly communicated as a function of career stage (e.g., pre-tenure), tenure track versus NTE status, maintenance of "research active" status and what that entails, etc. Beyond just providing guidelines about expected time commitment (e.g., 40:40:20 seems common for tenured, research active faculty at W&M), endeavor to articulate expectations qualitatively as well (e.g., contribution to culture and climate) and in terms of

outputs and outcomes (Benton, 2018), including learning outcomes beyond the classroom (e.g., from activities, events, etc.).

Sources of feedback

The problem: Teaching evaluations are often based on the narrow perspective of student surveys at the end of the semester. When peer feedback is solicited (from other educators), it is often just prior to tenure and promotion decisions and is treated as more of a “reality check” than as input to an ongoing evaluation and development process.

Recommendations: Evaluation and feedback should come from multiple sources. In addition to student feedback, the process could include periodic feedback from structured peer observations, for example (i.e., as part of regular annual evaluations, not just prior to tenure and promotion). Outside evaluators could also be used (e.g., evaluators from other departments, or even external consultants). In addition to feedback from peers, feedback should come from administrators or program leaders who are in a position to be aware of contributions outside the classroom (e.g., curriculum development, advising). Finally, the educators themselves should also provide feedback in the form of a self-assessment (e.g., a narrative about teaching philosophy, contributions, assessment of progress towards past goals, future goals, etc.).

Scope of measures

The problem: There is often an over-reliance on limited measures, with a bias towards quantitative data.

Recommendations: The process should incorporate multiple measures, both quantitative and qualitative, informal and formal. Two overarching concepts are to *triangulate* sources of information (find multiple sources at a point in time that corroborate or qualify a finding) and to look for patterns and change *over time*. There are a wide variety of possible types of measures and information beyond student ratings, including peer review of course materials, exit surveys and alumni ratings, review of teaching portfolios, etc. (see Berk, 2014 for details). It would not be possible to use all of these, of course. Instead, the process should be developed to incorporate different types of data that maximize the potential to both evaluate and develop great educators in a way that is sensitive to the resources required to do so.

Quality of student rating survey

The problem: Student surveys (aka Student Ratings of Instruction, or SRI) can vary quite a bit despite the existence of robust research on what factors should be included. However, possibly a bigger issue is the reliance on unidimensional, “global ratings,” such as “Overall, rate the effectiveness of the instructor” which is not diagnostic, and is relatively more subject to bias (Benton, 2018).

Recommendations: Eliminate unidimensional, “global” ratings – especially with respect to the instructor (vs. course). Instead, develop questions based on established, multi-dimensional conceptualizations of teaching effectiveness. For example, Marsh (1982) developed an evaluation of teaching survey based on nine dimensions that has been validated by substantial research. These nine dimensions are as follows:

- Learning/value
- Enthusiasm
- Organization

- Group interaction
- Individual rapport
- Breadth
- Examinations/grading
- Assignments
- Workload/difficulty

Marsh (1982) lists 35 multiple measures, termed “scale items,” which are needed to assess each dimension. Many W&M units rely on “global ratings” of overall effectiveness of: (1) instructor and (2) course. In Marsh’s (1982) framework, these ratings were actually considered scale (rather than dimension) items, and should not be used on their own, or as the most important of the ratings.

Use of statistics

The problem: Quantitative data (e.g., student ratings) are subject to misinterpretation and misuse. For example, typical practice is to report rating averages often without considering the rating distribution – or even the reliability of the data which depends on sufficient and representative samples. Similarly, there is a temptation to simply compare means between faculty (or between an individual faculty member and the global mean of faculty), without consideration of whether these mean differences are a) statistically significant (is the difference accounted for by measurement error?) or b) practically significant (what constitutes a “big” difference that warrants a different judgment of the educator including differences in rewards, such as merit pay?).

To complicate things, administrators or others who are in supervisory roles may not have the training needed to interpret statistics appropriately and fairly, or may not apply the same level of scrutiny that they would apply when reviewing academic research, for example, since “it is just a rating” (i.e., they may simply assume that all ratings are normally distributed, unbiased, etc.).

Recommendations: The process should address issues of:

- Reliability: Marsh (1982) suggested a minimum sample size of 10-15, which assumes a representative sample. The process should therefore explicitly address how to ensure a representative sample.
- Validity: Are we measuring what we think we are? Do the ratings provide information as intended? This depends on clear, unambiguous questions and sufficient time for evaluation.
- Fairness: What are we comparing the scores to? Similar courses in similar programs? How might the rating scores be biased?
- Consequences: How might the educator respond to the ratings? Might they “game” the process in a way that inflates ratings but hurts the learning experience?

For further details on the above, refer to Benton (2018), Stark and Freishtat (2014), and Marsh & Roche (1997).

Customization of approach

The problem: In an attempt to be fair and efficient, there is often a temptation to design a single process and then apply it consistently across all units (Schools, Areas, programs, etc.) and all individuals, regardless of job type, level of experience, etc. While this is

laudable, it can also be taken too far leading, paradoxically, to a system that feels unfair and that is less useful than one that is (somewhat) customized.

Recommendations: There may be elements of the process that can and should be designed centrally and adopted consistently across units (Schools, Areas, programs). This could include, for example, a consistent survey (Student Rating of Instruction). There may be other elements as well that could be developed centrally, but then customized locally, e.g., a universal structured “peer classroom observation” form. Consistency and efficiency are important. However, the process should then be customized within units as needed (e.g., possibly at the level of Schools) so that it can reflect the unique missions and goals of each school, as well as unique student populations (e.g., the undergrad A&S population is quite different than graduate students in the professional schools). In a similar sense, the process should also be customized to address different teaching populations (e.g., different career stages/ranks, NTE vs. TE, etc.).

Further, the process should address both teaching evaluation and development in balance. In other words, if the goal is to promote great teaching (and learning), then that must be explicitly manifest in the process, e.g., via faculty self-assessment and drafting of development plans, and/or the development of a peer-assisted teaching program that fosters a community of idea exchange and support (our Studio for Teaching and Learning Innovation provides a clear opportunity to foster the development of such a community, or pockets of teaching communities in various schools; see also Carbone et al. (2015) for a case study of a peer-assisted teaching program).

Finally, the process should be developed by those who will be involved (i.e., including administrators/faculty/students!).

MANAGEMENT AND EXECUTION

Rewards and Incentive for Positive Change

The problem: Currently W&M’s teaching evaluation system is primarily punitive and premised on a teaching as a fixed trait. Merit, retention, and compensation decisions frequently rely on overly-reductive SET scores. By the same token, awards that recognize teaching may also be influenced by a similar use of SET scores.

Recommendation: The redesign of W&M’s teaching evaluation should seek to reward faculty in their ongoing efforts to improve as educators. This should include a focused plan for improvement (Carbone 2015) and clearly stated goals and rewards for achieving stated teaching goals. Finally, any evaluative system that emphasizes growth in teaching should strongly consider using a peer-assisted development component as an integral part of the process.

Training for Evaluators

The problem: W&M’s current model of teaching evaluation does not include any training for evaluators (e.g., chairs, program directors, committee members) or for those being evaluated. This may be, in part, based on assumptions that all those who teach inherently understand what good teaching is. There are also assumptions that measures like SET can provide an accurate and objective measure of one’s teaching ability. Conversely, there is a lack of

understanding about the appropriate use of statistics, e.g., taking into account the statistical impact of low response rates.

Recommendation: Faculty who are responsible for evaluating their peers (including chairs and directors conducting merit evaluations) should be taught how to better understand SET scores from a statistical perspective, including for example, how to gauge the effect of response rates on SET scores. A significant portion of this training should include strategies for achieving and maintaining high response rates with pre-specified thresholds that guide the suitability of statistical interpretation.

The problem: SET scores (the primary measure of evaluation used by W&M) do not provide an accurate assessment of effective teaching. Instead, SET scores tend to measure qualities (like charisma, popularity, or likeability) that have little to do with effective teaching.

Recommendation: One possible solution would be to offer students more guidance as to the nature of the evaluation process. More specifically, W&M might consider helping students to reconceptualize evaluations as an opportunity to provide ongoing feedback for future development rather than a superficial numerical rating.

References for Literature Review Section

- American Sociological Association. (2019). "Statement on Student Evaluations of Teaching." https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_feb132020.pdf. Accessed 10/8/2020.
- Benton, S. L., & Young, S. (2018). "Best Practices in the Evaluation of Teaching. IDEA Paper #69." *IDEA Center, Inc.* <https://files.eric.ed.gov/fulltext/ED588352.pdf>. Accessed 10/8/2020.
- Berk, R. A. (2018). Start spreading the news: Use multiple sources of evidence to evaluate teaching. *The Journal of Faculty Development* 32(1), 73-81.
- Berk, R. A. (2016). Value of value-added models based on student outcomes to evaluate teaching. *The Journal of Faculty Development* 30(1), 73-82.
- Boysen, G. A., Kelly, T. J., Raesly, H. N., & Casner, R. W. (2014). The (mis) interpretation of teaching evaluations by college faculty and administrators. *Assessment & Evaluation in Higher Education* 39(6), 641-656.
- Carbone, A., Ross, B., Phelan, L., Lindsay, K., Drew, S., Stoney, S., & Cottman, C. (2015). Course evaluation matters: Improving students' learning experiences with a peer-assisted teaching programme. *Assessment & Evaluation in Higher Education* 40(2), 165-180.
- Carpenter, S. K., Witherby, A. E., & Tauber, S. K. (2020). On students' (mis) judgments of learning and teaching effectiveness. *Journal of Applied Research in Memory and Cognition* 9(2): 137-151.
- Flodén, J. (2017). The impact of student feedback on teaching in higher education. *Assessment & Evaluation in Higher Education* 42(7), 1054-1068.
- Fryberg, S. & E. Martínez, Eds. (2014). *The Truly Diverse Faculty: New Dialogues in American Higher Education*. Springer.
- Marsh, H. W. (1982). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology* 52(1), 77-95.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist* 52(11), 1187.

Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. DOI: 10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1.

SURVEY OF PRACTICES

The survey of practices yielded both quantitative and qualitative data for all five schools (SOB, SOE, SOL, SMS, A&S), 4 units within SMS, and 19 units within A&S. Summary tables for the quantitative data are provided in Appendix B. Raw responses from the qualitative questions are provided in Appendix C. Below, we've highlighted the major results, as we interpreted them.

PURPOSE AND PHILOSOPHY

Defining “effective teaching”

The survey attempted to understand if teaching expectations that defined effective teaching were provided across the schools and units. Twenty-nine units (including SOL, SOE, SMS, SOB) indicated that they provide concrete teaching expectations, although these expectations are primarily provided in the form of number of teaching credits or number of classes. Four units (all A&S) indicated that they provide “vague” teaching expectations, and one unit (A&S) noted that “no expectations” are communicated at all. Definitions of effective teaching were highly variable across units, with a handful of units applying multiple, explicit metrics (i.e., adequate preparation, regular class meetings, regular evaluation of student work, organized and articulated lectures, innovativeness, timeliness, availability, etc.) to others relying solely on number of credit hours or students taught.

In many units, teaching expectations are communicated “verbally” or are simply “understood,” as unit culture. A handful of units actually provide written teaching expectations via annual emails and written unit policies. Often, teaching expectations are communicated one-on-one (from chair or program director to faculty member), rather than to an entire unit.

Purpose of the Review Process

The survey attempted to answer the question of the effectiveness of teaching evaluation in meeting certain goals including *Faculty evaluation*, *Faculty development*, *Curriculum development*, and *External accreditation*.

QUESTION: For each of the goals listed below, rate the extent to which your unit's approach to evaluating teaching annually is effective in accomplishing it. In the table below, numbers indicate the percent of units within A&S and SMS that report that their approach to annual evaluation is Always or Almost Always effective in accomplishing that goal. Yes/No indicates whether the Dean (or Associate Dean) of the SOB, SOE, and SOL reports that their approach to annual evaluation in that school is Always or Almost Always effective in accomplishing that goal. Gray shading highlights a “No” school response or below 50% unit response. For more explanation, see Appendix B.

	A&S	SOB	SOE	SOL	SMS
Faculty evaluation (e.g., annual merit reviews)	20.8	Yes	Yes	No	100
External accreditation	18.8	No	No	No	0
Faculty development (i.e., ongoing improvement in teaching ability and performance)	10.0	No	No	No	0

Curriculum development (i.e., ongoing improvement of courses and programs)	4.8	Yes	No	No	0
--	-----	-----	----	----	---

While some schools (SOB, SOE, SMS) report that their approach to evaluating annual teaching is always or almost always effective in faculty annual merit review, only 21% of A&S units agree. All of the schools (and 90% of A&S units) report that their approach to evaluating annual teaching fails to accomplish faculty development (i.e., the ongoing improvement in teaching ability and performance)—despite the emphasis placed on this particular goal in the literature. Similarly, none of the schools (and only 19% of A&S units) report that annual teaching evaluations are effective for external accreditation. The SOB and 5% of A&S units suggest that annual evaluation of teaching is successful in improving courses and programs (=curriculum development).

Scope of Review Process

The survey attempted to answer the question of what other factors (beyond classroom performance) were considered when a unit evaluates teaching, including *Course development (i.e. individual courses)*, *Curriculum development (i.e. new or revised programs, minors, majors)*, *Teaching innovation*, *Individual student support (i.e. office hours, mentoring)*, *Promoting diversity and inclusion in teaching*, and *Course materials*.

QUESTION: Beyond classroom performance, to what extent are each of these other activities included when your unit evaluates teaching annually? In the table below, numbers indicate the percent of units within A&S and SMS that report that these other activities are Always or Almost Always included when their unit evaluates teaching annually. Yes/No indicates whether the Dean (or Associate Dean) of the SOB, SOE, and SOL reports that these other activities are Always or Almost Always included when their school evaluates teaching annually. Gray shading highlights a “No” school response or below 50% unit response. For more explanation, see Appendix B.

	A&S	SOB	SOE	SOL	SMS
Course development (i.e., individual courses)	69.2	Yes	Yes	Yes	75
Curriculum development (i.e., new or revised programs, minors, majors)	46.2	Yes	No	No	75
Individual student support (i.e., office hours, mentoring)	46.2	No	Yes	Yes	50
Teaching innovation (i.e., new approaches, methods, technologies)	46.2	Yes	Yes	Yes	25
Promoting diversity and inclusion in teaching	11.5	No	No	Yes	25
Course materials	12.0	Yes	No	No	0
Other (please specify)	100	No	No	No	0

The evaluation of teaching effectiveness at W&M is narrowly defined relative to effective practices outlined in the literature. At W&M, evaluation is based almost exclusively on classroom performance and course development. 69% of A&S units, as well as 75% of SMS units, SOB, SOE, and SOL indicated that *Course Development* was “Always” or “Almost

Always” included. Other teaching activities, including individual student support and teaching innovation are used by more than half of the units at W&M. The next most likely factors to be considered included *Curriculum Development* (46% of A&S units, 75% of SMS units, and SOB), *Individual Student Support* (46% of A&S units, 50% of SMS units, SOE, SOL) and *Teaching Innovation* (46% of A&S units, 25% of SMS units, SOB, SOE, SOL). The least likely factors to be considered when evaluating teaching included *Course Materials* (only 12% of A&S units, plus SOB) and *Promoting Diversity and Inclusion* (11.5% of A&S, 25% of SMS units, SOL). This is despite the fact that “belonging” and embracing diversity is one of our core values.

APPROACH

Sources of information

The survey attempted to answer the question to what extent particular sources of information were used to evaluate teaching including *Student course evaluations*, *Peer classroom observation* (i.e. by colleagues), *Self-evaluation* (i.e. by oneself), *Assessment of course materials* (i.e. syllabus, assignments), *Assessment of student work*, *Student grade distributions*, and *Quantity of teaching*.

QUESTION: For each category listed below, rate the extent to which this information is used when your unit evaluates teaching annually. In the table below, numbers indicate the percent of units within A&S and SMS that report that this information is Always or Almost Always used when their unit evaluates teaching annually. Yes/No indicates whether the Dean (or Associate Dean) of the SOB, SOE, and SOL reports that this information is Always or Almost Always used when their school evaluates teaching annually. Gray shading highlights a “No” school response or below 50% unit response. For more explanation, see Appendix B.

	A&S	SOB	SOE	SOL	SMS
Student course evaluations	92.3	Yes	Yes	Yes	100
Quantity of teaching (e.g., course load, credit hours, contact hours, etc.)	61.5	No	Yes	Yes	100
Assessment of course materials (i.e., syllabus, assignments)	34.6	Yes	Yes	Yes	0
Self-evaluation (i.e., by oneself)	23.1	No	Yes	Yes	75
Other (please describe)	100	No	No	No	0
Peer classroom observation (i.e., by colleagues)	19.2	No	No	No	0
Student grade distributions	15.4	No	No	No	0
Assessment of student work	7.7	No	No	No	0

The vast majority of schools and units at W&M rely on three approaches to evaluating teaching: SETs (discussed above), quantity of teaching (i.e., course load, credit hours), and assessment of course materials (i.e., syllabi, assignments). Quantity of teaching is always or almost always considered in annual merit in the SOE, SOL, all SMS units, and 62% of A&S units. Assessment of course materials is always or almost always used in the SOB, SOE, SOL, 35% of A&S units, but not SMS.

Other approaches to evaluating teaching that are championed in the published literature are rarely used at W&M for annual merit. For example, self-evaluation, which is often recommended particularly for faculty development, is always or almost always used by the SOE, SOL, 75% of SMS units, and 23% of A&S units, but not by other schools. Peer classroom observation, often listed as one of the most effective strategies for teaching evaluation, is always or almost always used by only 19% of A&S units, and not by other schools. Assessment of student work, which provides a benchmark for learning, is always or almost always used by only 7.7% of A&S units and not by other schools.

Scope of Measures

The survey attempted to answer the question to what extent certain measures are used to evaluate teaching including *Student scores, Student comments and other (please describe)*.

QUESTION: For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching annually. In the table below, numbers indicate the percent of units within A&S and SMS that report that this information is Always or Almost Always used when their unit evaluates teaching annually. Yes/No indicates whether the Dean (or Associate Dean) of the SOB, SOE, and SOL reports that this information is Always or Almost Always used when their school evaluates teaching annually. Gray shading highlights a “No” school response or below 50% unit response. For more explanation, see Appendix B.

	A&S	SOB	SOE	SOL	SMS
Student evaluation scores	84.6	Yes	Yes	Yes	100
Student evaluation comments	61.5	No	Yes	Yes	75

All schools and units (with the exception of two A&S units) report that they always or almost always use SETs to evaluate teaching for annual merit. How these evaluations are used varies by school and unit. While all schools report that they always or almost always use the numerical scores from SETs, 15% of A&S units report that they do not. Student comments are always or almost always used by fewer units—61.5% of A&S and 75% of SMS units. While SOE always or almost always uses SET comments, SOB does not.

Quality of Student Rating

The survey attempted to answer the question to what extent certain factors are used to evaluate teaching including *How student evaluation scores **change over time**, How student evaluation scores differ by **course type**, How student evaluation scores are effected by **faculty gender**, How student evaluation scores are affected by **faculty race or ethnicity**, How student evaluation scores are affected by **faculty sexual orientation, gender identity, or gender expression**, How student evaluation scores are affected by **expected grade**.*

QUESTION: For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching annually. In the table below, numbers indicate the percent of units within A&S and SMS that report that this information is Always or Almost Always used when their unit evaluates teaching annually. Yes/No indicates whether the Dean (or Associate Dean) of the SOB, SOE, and SOL reports that this information is Always or Almost Always used

when their school evaluates teaching annually. Gray shading highlights a “No” school response or below 50% unit response. For more explanation, see Appendix B.

	A&S	SOB	SOE	SOL	SMS
How student evaluation scores differ by course size	46.2	Yes	No	Yes	50
How student evaluation scores differ by course type	46.2	No	No	Yes	50
How student evaluation scores change over time	23.1	No	Yes	Yes	25
How student evaluation scores are affected by faculty gender	3.8	No	No	Yes	25
How student evaluation scores are affected by faculty race or ethnicity	3.8	No	No	Yes	25
How student evaluation scores are affected by expected grade	3.8	No	No	No	0
How student evaluation scores are affected by faculty sexual orientation, gender identity, or expression	3.8	No	No	No	0

Although tracking of SET scores over time can be helpful for faculty development, relatively few W&M units report that they always or almost always look at how scores shift over time for faculty (23% of A&S, 25% of SMS, SOL, SOE) for annual merit. In contrast, the effects of course size and course type on SET scores are always or almost always considered by many more units (46% of A&S, 50% of SMS, SOL, SOB (course size only)). Across all of the schools and units, only one A&S unit always or almost always takes into consideration how SET scores may be affected by expected grade.

Digging even more deeply, the majority of W&M schools and units report that they do not always or almost always consider the explicit effect that gender, race/ethnicity, or gender/sexual identity bias may have on SET scores. Only SOL, one SMS unit, and one A&S unit report that they always or almost always consider the inherent bias in these scores by gender and race.

Use of Statistics

The survey attempted to answer the question to what extent certain statistical data were used in teaching evaluations including *Mean student evaluation scores*, *Median student evaluation scores*, *Standard deviation of student evaluation scores*, *Response rate of student evaluations (%)*, and *Course size*.

QUESTION: For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching annually. In the table below, numbers indicate the percent of units within A&S and SMS that report that this information is Always or Almost Always used when their unit evaluates teaching annually. Yes/No indicates whether the Dean (or Associate Dean) of the SOB, SOE, and SOL reports that this information is Always or Almost Always used

when their school evaluates teaching annually. Gray shading highlights a “No” school response or below 50% unit response. For more explanation, see Appendix B.

	A&S	SOB	SOE	SOL	SMS
Mean student evaluation scores	88.5	Yes	Yes	Yes	100
Course size	45.5	Yes	No	Yes	100
Response rate of student evaluations (%)	38.5	No	No	Yes	66.7
Median student evaluation scores	28	No	No	Yes	66.7
Standard deviation of student evaluation scores	15.4	No	No	No	66.7

Units are still relying, always or almost always, on mean SET scores (89% of A&S and 100% of SMS units, as well as SOB, SOE, and SOL) for annual merit, rather than other metrics that are recommended in the published literature including median or standard deviation. The only units that report always or almost always using standard deviation (or some other measure of variation) are three SMS and four A&S units. Similarly, few units (38.5% of A&S, 67% of SMS, and SOL) report that they always or almost always use response rate in their assessments, despite the fact that sample size issues are known to greatly affect SET scores.

Customization of approach

Participants did not indicate that additional or different approaches were used with evaluating teaching for tenure-eligible versus non-tenure-eligible faculty. However, half of participants did indicate that units applied additional or different approaches when evaluating teaching for pre-versus post-tenure faculty, for example, more “classroom visitations”, or “a focus on improvement” in the case of pre-tenure faculty. The other half noted that there was not a difference in how teaching was evaluated when comparing pre vs post-tenure faculty.

MANAGEMENT & EXECUTION

Rewards and Incentive for Positive Change

When examining how units recognize or reward improvement in teaching, the participants indicated that this is done in the form of higher merit scores, which typically equates to salary raises and university nominations for teaching awards. However, four respondents noted that teaching improvement was not recognized or rewarded.

Training for Evaluators

Participants indicated that mentoring may be used, which is informal and typically for pre-tenure and new faculty or on an as needed basis. However, specific training for evaluation was not indicated.

REVIEW OF POLICIES

Policies for the Evaluation of Teaching

All units of the College are bound by the certain statements in the Faculty Handbook regarding the evaluation of teaching: One of the criteria for retention, promotion and tenure is “*conscientious and effective teaching with proper command of the material of [the faculty members’] fields, and helpfulness to their students.*” (III.C.1), and reviews for retention, promotion, and tenure “*shall be conducted in accordance with the general categories delineated in III.C.1. above, with the procedures delineated in III.C.1.b., and with the standards and procedures adopted by the faculty member’s program, department, and/or school. The information considered shall include the candidate’s curriculum vitae, self-evaluation, and student evaluations, as well as some evaluation of the candidate’s teaching based on at least one method other than student evaluation*” (III.C.1.b.i; = III.C.1.b.ii.B; III.C.1.b.iii).

This sets some basic guidelines that all units follow, but the details lie in the “*standards and procedures adopted by the faculty member’s program, department, and/or school,*” and there is wide variation in the policies of the different units. What follows is an overview of those policies. As these are personnel policies designed for the process of determining whether faculty have met certain benchmarks for tenure, promotion, etc., they tend to focus on summative rather than formative evaluation, although most do rely on information beyond SET scores. Our survey indicates that a handful of units employ assessment of teaching for formative purposes, but those activities largely lie outside the realm of official personnel policy.

I. School of Business: Teaching is reviewed annually for all faculty; SETs are the primary input for these reviews. These are supplemented by optional information provided by the faculty member about course innovations or new course development. Retention, tenure, and promotion reviews, as well as periodic post-tenure reviews use the following evaluation methods:

- A. Peer evaluation: Three faculty conduct at least two classroom visits and complete an evaluation form for each visit. The evaluation includes a review of course materials, a pre-visit meeting with the instructor, an in-class evaluation, and follow up with the instructor at the discretion of the reviewing faculty. Members of the Personnel Committee that vote on retention/tenure/promotion may also conduct classroom visits.
- B. SET: School-approved questionnaires are distributed in all degree credit classes. Faculty summarize student course evaluation results on a spreadsheet. The Personnel Committee is not explicitly guided to emphasize any single item of the student questionnaires; special attention is given to written comments.
- C. Other evidence of teaching effectiveness may be submitted.

II. School of Education: Faculty submit a self-assessment of teaching performance in their annual review materials. The following criteria are used for tenure/promotion with respect to teaching:

- A. Philosophy of teaching statement
- B. Statement of teaching, supervising, and advising responsibilities
- C. Original copies of end-of-course SETs for all courses taught
- D. Other teaching evidence, including:
 - 1. Evidence of content expertise (e.g., training, professional development)
 - 2. Evidence of course planning and instructional design

3. Evidence of quality in instructional delivery (e.g., course ratings at or above SoE criteria, use of assessment feedback for instructional improvement, peer evaluation, mid-course feedback)
 4. Indicators of assessment and evaluation practices (e.g., assessment of course objectives, evidence of student learning or outcomes)
 5. Description of other teaching responsibilities (e.g., thesis/dissertation direction, independent study, field supervision, and program advising)
- E. Self-evaluation of teaching
- F. Future plans for teaching
- G. For promotion to full professor, significant contributions to program development and improvements

III. School of Law: All full-time faculty members must submit an annual report to the Dean that describes and evaluates their teaching activities for the past year. The faculty member must include number of credits and students taught, as well as other teaching activities over the evaluative period. The Vice Dean or Associate Dean provides information on the SETs for the faculty member prior to a meeting with the Dean, Vice or Associate Dean, and the faculty member. Any teaching issues are discussed in the annual merit review meeting.

The SOL's Procedures for Retention, Promotion and Tenure require evaluation of untenured faculty members' courses. In addition to reviewing the junior faculty member's SETs, the tenured and, in the case of some junior NTEs, NTE faculty with security of position may attend the junior faculty member's classes at designated time periods set by the SOL's Status Committee. During an RPT review period, the Status Committee also invites comments from students. All junior faculty are assigned faculty mentors to guide their development as educators and scholars (when applicable).

IV. School of Marine Science: Faculty summarize their teaching activities in their annual review and retention/tenure/promotion materials. Policies for annual merit evaluations and retention/tenure/promotion both state that "... *faculty and student evaluations of the effectiveness of teaching and advising will be considered.*" This is accomplished through SETs for all courses with the exception of undergraduate, thesis, and dissertation research credits. The Associate Dean of Academic Studies (AD-AS) reviews SET results and communicates them to the department chair for annual merit reviews, or in a formal letter to the tenure and review committee for retention/tenure/promotion cases. The AD-AS also comments on the quality of a faculty member's academic advising, research training, and facilitation of student professional development. SET results for all courses with more than three students are provided to the instructor. Results for courses with three or fewer students are not provided to instructors to protect confidentiality.

V. School of Arts & Sciences: Arts & Sciences is the largest unit, and comprises numerous diverse departments and programs, so its policies on the evaluation of teaching are not so easily summarized. Some unit-wide policies are contained in the Dean's memo on "*Arts & Sciences Procedures on Tenure, Promotion, and Interim Review Processes*," last updated May 5, 2017 (with a small footnote added in response to the coronavirus crisis of Spring 2020). Here are the

sections of the memo dealing with the evaluation of teaching in the various review processes of a candidate's career (I.C.12-14):

12. A description of the kinds of courses offered by the candidate, such as survey or introductory, upper level, or seminar must be included in the report of the Chair, the department, or the departmental personnel committee....

13. A second means of evaluating the faculty member's teaching must be included besides student evaluations.... (e.g., review of exams and syllabi, peer observation)

14. Candidates for tenure will include all evaluations for each course taught since they arrived at the College and those for promotion will include all student evaluations for each course taught since tenure or for the previous eight consecutive years, whichever is shorter.¹ The department is required to provide a single table summarizing the candidate's scores for all courses on the question "what is this instructor's overall teaching effectiveness." The department should also include in this table a specified numerical comparison (e.g. the departmental mean) to other departmental courses. Additionally, the department must include the comments from the student evaluations in one of the following forms: a PDF of the complete evaluations or a compilation of all student comments, clearly identified by course, semester, and year.

As with the language in the Faculty Handbook and the Provost's memo, the Dean's memo provides some general guidelines that departments and programs tend to follow, but a complicating factor is that the Faculty Handbook specifies that for Arts & Sciences faculty, it is not Arts & Sciences policy but department and program policy that should be followed in the evaluation of faculty (III.C.b.i). This means that in cases where the duly approved policies and procedures of a department or program are at variance with the guidelines in the Dean's memo, the former must take precedence.

All the subdivisions of Arts & Sciences follow the general directive of the Faculty Handbook that teaching is to be evaluated by means of student questionnaires and by at least one other method for tenure and promotion, but there is great diversity from one department or program to the next as to a) how student questionnaires are to be used and interpreted, b) what other means of evaluation are to be employed, and c) whether or not the same means and materials are used to evaluate faculty at all ranks and categories (TE/NTE/adjunct, etc.) and in all evaluative contexts (annual merit, interim review, tenure, promotion, etc.). The following is a summary of that diversity, based on a collection of policies maintained by the office of the Dean of A&S. This collection represents 23 of the 50+ departments and programs in A&S² – it is to be noted that not even the Dean's office was able to supply ready access to the personnel policies of every unit. It is also to be noted that some policies have not been updated in more than a decade and are clearly out-of-date in some respects; e.g., mandating the ways in which paper evaluation sheets must be administered and collected.

SETs: while all departments and programs require the use of SETs, there are few commonalities between the units on how they are to be used and interpreted.

- Many departments and programs (9 of 23) specify that students' written comments must be considered in addition to the numerical scores.
- Some (6 of 23) specify which questions on the questionnaire are to be given primary attention, but these questions are often identified vaguely and/or inconsistently; examples include "[questions] rating the instructor and the course," "categories rating the instructor and the course," "overall teaching effectiveness," "overall quality of the course," "instructor's overall teaching," "overall course evaluation," "how would you rate your instructor," and "learn anything."
- Some (5 of 23) specify that an individual's scores should be compared to a departmental or program mean, either for all courses or for courses of a similar type and/or enrollment size.
- A few (3 of 23) comment explicitly on potential biases and limitations of the numerical scores, including the problems of small sample sizes and low response rates.

Other methods: All departments and programs state that at least one other means of evaluation must be used, but again, there is considerable variation from unit to unit.

- Almost all (19 of 23) explicitly include syllabi among the things that can or must be submitted for evaluation.
- Exams can or must be evaluated by a narrow majority of units (12 of 23).
- Classroom visits by peers or senior colleagues are mentioned by a solid majority (15 of 23), though only a few (4 of 23) specifically require such visits as part of the evaluation process.
- Other methods frequently mentioned include assignments (8 of 23), grades (5 of 23), and generic "course materials" (9 of 23)
- More rare (specified by only one or two units, and generally only in a mixed bag of possible means of evaluation) are such things as samples of student work (1), unsolicited student comments (1), the results of department/program-level surveys of graduating concentrators and/or alumni (5), and letters of support from faculty within or outside of W&M (2)
- A few units (5) state explicitly that either the faculty member or the person(s) evaluating him/her may include other sorts of material that they deem relevant.

Other Considerations: In addition to the above, many departments and programs specify other pedagogical activities that can or should be taken into consideration in evaluations, including the following (in descending order of frequency of mention; the numbers represent the number of units out of the 23 studied that mention each item explicitly):

Specified by a slight majority of departments and programs:

- Mentoring theses, dissertations, and other student research efforts (13)
- Overload teaching, including independent studies (12)

Specified by several but less than a majority:

- Creating new courses (11)
- Advising (concentration and pre-major) (6)
- Course size (4)
- Level and type of courses taught, including whether courses are writing-intensive (4)
- Updating existing courses (4)
- Engaging in curricular development (4)
- Attending professional development workshops, seminars, etc. (4)

- Making one's self available to students (3)
- Giving guest lectures in colleagues' classes (3)
- Serving on honors and examination committees (3)

Specified by one or two:

- Course load (2)
- Using clear and appropriate testing and grading criteria (2)
- Leading field trips (2)
- Coaching/leading teams and groups for performances and academic competitions (2)
- Summer teaching and study-abroad (2)
- Winning grants for teaching development (2)
- Winning teaching awards (2)
- Encouraging discussion in appropriate classes (1)
- Writing and administering comprehensive exams (1)
- Organizing large events (1)

Levels of evaluation: Ten of the 23 units specify different procedures and requirements either for different levels of assessment (annual merit review, tenure, promotion to full, etc.) or for different categories of faculty (TE, NTE, adjunct, etc.). Here are some examples:

- One unit has a sizable list of measures that can be applied in addition to SETs for annual merit, but for tenure and promotion it is specified that at least one of the following must be used: a) two classroom observations; b) review of syllabi & exams; c) a *“teaching portfolio”*. For NTEs, syllabi are specified in addition to one other means chosen by the candidate.
- One unit has different requirements for faculty based on length of service
 - Faculty with 1-3 years' service must, each semester, submit a) a self-evaluation; b) SETs for all classes c) course syllabi and *“information sheets”*, and d) confer with the chair. Class observations are conducted by the chair *“as deemed necessary”*.
 - T/TE faculty with 4+ years' service must confer with the chair at least once every year.
 - *“not promotion eligible”* faculty with 4+ years' service must confer with the chair at least once a year and submit SETs for all courses and sections with the exception of one-credit courses, for which only one section once a year is required.
- A handful of units require a report from faculty visitations at various levels. For one, the visits are arranged by the chair and are not specified as to number and frequency. For another, yearly observations by two tenured faculty are required at the pre-tenure level, and for promotion to full, two observations carried out within two years of the beginning of candidacy. One unit requires that new faculty visit a senior faculty's course and have their own courses visited by a senior faculty member within their first semester at W&M.
- One unit requires class observations and grade distributions for annual merit, but does not specify them for tenure/promotion; for NTEs submission of anything besides teaching evaluations is optional
- One unit requires grades for annual NTE merit evaluation, and yearly class visitations for NTE and pre-tenure faculty.

As mentioned above, a few of the departments and programs (6 of 23) present directives for formative use of the evaluation process in their official personnel policies. Here are some examples:

- Some units require one-on-one faculty conferences with the chair on a yearly basis (and one unit does so on a semesterly basis for newer faculty. In one unit, the personnel committee is to conduct these conferences, which are recommended for pre-tenure faculty and available upon request to post-tenure faculty. The personnel committee is also charged with arranging class observations upon request.
- In addition to yearly conferences with the chair, one unit also recommends that instructors use SETs to improve their teaching as well as consider peer observation and video-recording of class as means to improve teaching. It also specifies that the department or program should seek to provide opportunities for instructors to improve their teaching through encouraging team-teaching and sponsoring group discussions on pedagogical issues.
- One unit has a system whereby new faculty are required to visit a senior faculty's course and have their own course visited by a senior faculty member within the first semester. Following both visitations the new faculty member meets with the visitor/visitee faculty to discuss their observations.

As stated previously, our survey indicates that many other departments and programs have procedures in place to make formative use of teaching evaluations, but have not opted to make these procedures part of their official personnel policy.

Policy vs. Literature Review:

The literature review highlighted a number of best practices for teaching evaluation and made a series of recommendations for implementation at W&M. Here we compare these recommendations to current policies across the university in a number of areas:

Defining effective teaching: The literature review highlighted the importance of defining effective teaching, setting clear expectations for teaching, and defining the type of activities that constitute teaching. Current policies across programs, departments, and schools generally do not offer a definition of effective teaching, so this may be a key area for W&M or School-wide focus in the coming year. Policies in the various units do frequently recognize the wide variety of activities beyond traditional classroom teaching that may be considered when evaluating the faculty (e.g., curriculum development, advising, and mentoring), although there tends to be great variability across the units in the types of activities that are considered. While some variability is to be expected given differences among the disciplines, there may be merit in adopting more consistent language across the units to embrace a wide view of what constitutes teaching.

Formative vs. summative evaluation: The literature review also highlighted that teaching evaluations tend to be primarily summative for the purposes of determining annual merit scores or setting benchmarks for tenure and promotion. Best practice, however, suggests that evaluations should be formative with the stated goal of helping educators improve over time (i.e., a “growth mindset”). As summarized below, a number of the policies across W&M units do encourage use of evaluation tools that could be used in a formative way, and some even present directives for formative use of the evaluation process. The policies themselves however tend to focus on a summative approach to evaluation for the purposes of merit and tenure/promotion. The existing framework does offer the potential for transitioning to a more formative approach.

Multiple modes of evaluation: Best practices in teaching evaluation emphasize the use of multiple modes of evaluation that go beyond SETs. These should include both quantitative and qualitative, informal and formal metrics. Overall policies (Faculty Handbook, A&S Dean's Memo) embrace this approach by calling for at least one metric in addition to SETs. Policies within specific units also tend to embrace this practice. All A&S policies surveyed state that at least one other means of evaluation beyond SETs must be used (e.g., peer observation, review of course materials, development of new courses or enhancement of existing courses, mentoring, advising, professional development), and three of the four professional/graduate school policies also call for additional approaches. However, there is considerable variation from unit to unit in the types of additional evaluation tools to be used, and constraining reviews to a single additional approach may limit our ability to take a formative approach.

SETs: The literature review emphasized the importance of properly using SETs, including elimination of unidimensional, "global" ratings (e.g., "rate the overall effectiveness of the instructor"). The review also highlighted numerous issues with statistical summaries of these evaluations. While all unit policies require the use of SETs, there is great diversity among A&S units as to how student questionnaires are to be used and interpreted, and limited information on how they are to be used in the professional/graduate schools. Differences in A&S include the degree to which (1) student comments are to be used, (2) emphasis is placed on particular questions, and (3) results are statistically evaluated. A&S units are required to summarize scores for the instructor's overall teaching effectiveness, and include a numerical comparison of mean scores, which are contrary to best practices identified in the literature review.

Timing of evaluations: The literature review recommends that in-depth teaching evaluations (i.e., beyond SETs) should be conducted more frequently than tenure and promotion cases. Again, there appears to be wide variation in W&M policies on this. A&S policies call for faculty to be evaluated annually based on SETs and at least one other metric. A minority of A&S units specify different procedures and requirements for tenure/promotion. Professional/graduate school policies rely on SETs for annual merit reviews, and more in-depth evaluations only for tenure/promotion with the exception of SOB which also requires in-depth, periodic, post-tenure reviews.

Customization: The literature review highlighted the importance of striking a balance between university-wide policies and unit-specific policies. W&M appears to have developed such a balance, with limited university-wide policies defined in the Faculty Handbook, and school- and unit-specific policies based on individual disciplines. While this is commendable, it has also led to a large degree of variability in how teaching is evaluated across units, including some discrepancies between A&S-wide and unit-specific policies. The various units have also in some but not all cases specified different procedures and requirements for different categories of faculty (TTE, NTE, adjunct).

Policy vs. Practice:

Overall, the results of our survey show that the practice in most units, as reported by the survey respondents, is generally consistent with the official policies. Where there are discrepancies, they

most often consist of units employing more precise procedures and standards than are spelled out in the policies. For instance, in response to the question, **“Are some questions on student evaluations forms considered more important to your unit than others?”** the vast majority of respondents said that in practice the two final questions on the standard evaluation form, on the overall quality of the course and the overall quality of the instructor, are given the most weight, but that is not communicated clearly in the official policies of most units, which are generally much more vague on this issue. To the question of **“Does your unit set a target minimum response rate per course for SETs?”** most respondents answered ‘no’, which accurately reflects the absence of any such standard in most stated policies. One respondent, however, reported that while the policies of their unit specify a preferred minimum response rate, that response rate has been mostly ignored in recent years as it is a relic of the days when paper evaluations were filled out in the classroom. As is the case for many units, the policies have not been updated for the era of electronic evaluations. As to **“how are expectations communicated?”** the survey responses accurately reflect the absence of any clear guidance in most of the policy documents: *“This is done in conversation. There is seldom a written record;”* *“Through the chair; mostly through dept. culture and meetings;”* *“no formal process for discussions about teaching to take place;”* *“word of mouth.”*

On the issue of formative assessment, many respondents to the survey reported that their units do more to apply the results of teaching evaluations to help faculty improve their teaching than is reflected in the official policies. The head of one of the graduate/professional schools sets a high standard, saying *“I ... prepare individual statements for each faculty member concerning the positive and negative feedback that I glean from student evaluations of their courses If applicable, I point out areas of improvement so that faculty can focus their efforts accordingly. All of this is received in written form (memo) and also discussed verbally during the annual performance evaluation.”* In many other units the process is more informal and reliant on the ability of the chair or other authority figure to recognize a problem and take action: *“The chair's narrative assessment provides such feedback. It is typically quite brief, but might highlight recurring themes in the written comments from students”;* *“The Chair disseminates their merit score for teaching and if the score suggests there are major problems, the Chair meets with the faculty member.”* *“Depends on the chair! And whether the faculty member is new or up for evaluation.”*

Many respondents state that there is a regular mentoring process in their unit, particularly for pre-tenure faculty, e.g., *“NTE faculty with teaching evaluations below 3.8 average teaching effectiveness score are mentored by colleagues. Junior faculty with low teaching evaluation scores are mentored by colleagues.”* Similarly, more respondents report that class visitations occur regularly in their unit than is reflected in the policies.

In all, the survey suggests that units tend to do more to evaluate teaching and to assist faculty in improving their teaching than they are required to do by their personnel policies. In many ways this is a good thing. The lack of detailed guidance from the policies leaves units the freedom to be responsive to the diverse circumstances that faculty find themselves in. At the same time there is a price to be paid for too much freedom. As one respondent notes, *“There are many things that are agreed upon but not written into policy, and this creates bad feelings, misunderstandings, and faculty abuse.”*

CASE STUDIES FROM OTHER INSTITUTIONS

In order to understand how other higher education institutions are evaluating teaching, we conducted a case study approach, by reaching out to eight universities (Smith College, University of Kansas, University of Oregon, University of Southern California, and Vanderbilt University). This sample includes both public and private, and small and large institutions, across different geographic regions.

What were the goals/objectives of the review/implementation? Including working definitions of excellence in teaching

The goals of Kansas, Oregon, USC, all expressly addressed a concern with clearly defining excellence in teaching; Smith's and Vanderbilt's objectives were more directly focused on improving the evaluation process, making it more transparent, and making it comply with current research on best practices in evaluation – in effect, all 5 universities share these last concerns, whether for the sake of better student input, equity in promotion, merit, and tenure, or to expand student participation in the face of declining rates of participation.

What were the most useful resources (on or off campus) identified by the reviewers?

Resources vary quite a bit depending on the kind of institution (Smith and USC and Oregon have little in common with respect to size, demography, or funds). All recognize the importance of some kind of “Center for Excellence in Teaching.” USC made a substantial commitment to instructional design, hiring two, and later three full-time staff to address evaluation across their vast university. Kansas received an NSF grant as part of a wider group to improve STEM evaluation. In general, it seems both funds and an administrative center are vital to reforming and implementing changes in teaching evaluation.

What were the major recommendations of the review/implementation?

The recommendations generally follow from the goals of the review. Kansas and Vanderbilt, in particular, hoped to increase student participation in the evaluation process either via easily accessible technology, reforms in the evaluation questions, or simply providing more time for student input. Smith stressed the importance of including students in the process of revising and implementing any changes to evaluation. Both USC and Oregon made substantial efforts to make peer and materials evaluation more robust. USC emphasizes student feedback in the form of comments, but wants to significantly limit the impact of SET scores in evaluating teaching.

How were these recommendations communicated to administrators, faculty, and students?

Communication was considered a key factor throughout the process of researching and ultimately changing the teaching evaluation process. All of the institutions that were surveyed emphasized the need to “over-communicate.” In particular, Oregon utilized the members of the task force itself to ensure that accurate information was being spread to the university. They met with every academic department and a variety of student groups to answer questions and garner

feedback from these constituencies. Vanderbilt held a series of voluntary focus groups where students, faculty, and staff were able to learn more about the upcoming changes. Institutions also utilized university wide emails and creating webpages detailing changes made.

What do you wish you had done differently with respect to the review/implementation?

Most of the institutions surveyed either had not completed the review or implementation or did not have any additional feedback. Vanderbilt did share that it was imperative to figure out how to convey the importance of changing the evaluation process to faculty members. They articulated that without faculty buy in from an early stage, it would be challenging to enact change at a systemic level. The idea of faculty buy-in was echoed from all of the other institutions when discussing challenges and additional questions.

What do you wish you knew at the beginning of the review/implementation-- that you know now?

In most cases institutions overall had felt that they had made changes that were congruent with the needs and culture of the institution itself. Vanderbilt and Smith discussed the need for including students in the process of reviewing and implementing changes. Students are an important part of the teaching evaluation process, so getting their feedback is critical. Oregon discussed that they wished they had known how important it was to fully understand what the institution was evaluating. This process made Oregon set specific goals and standards for teaching at their institution. Additionally, Smith highlighted partnership with Institutional Research as early as possible in the process and thinking critically about how to present information to the different audiences at the university. Lastly, Vanderbilt recommended utilizing a data warehouse in order to be able to analyze the data collected.

Case Studies Compared with WM Practice Survey

In the five cases examined (Kansas, Oregon, Smith, USC, and Vanderbilt) it is clear that the reviews carried out by these institutions identified some common themes. Participation in SETs were declining, especially once the surveys went online; the feedback generated by students tended to be vague, or otherwise unhelpful; student feedback tended to be summative and quantitative rather than formative and qualitative; the sources of evaluation tended to be quite limited, with no real role for faculty or peer evaluation. Each of these reflect the problems that our Practices Survey also identified at William & Mary. Most faculty in most units are not given any feedback on how to improve their teaching by their annual review process (and most didn't have access to support until STLI was established); quantitative scores from students tend to dominate these evaluations, even as participation in these evaluations has declined; and there are very few clear definitions of what constitutes excellent teaching by department, program, or discipline that are given to either faculty or students.

Comparing the case studies to the Survey evidence suggests a few "best practices" that might be adopted. First and foremost, it is important to offer clear and consistent definitions of "Excellence in Teaching." The case studies suggest that this is ideally both a centralized and "local" practice; that is, broad expectations for (e.g.) innovative, inclusive, engaged, and research-centered teaching can be formulated for the College as a whole; and individual units,

departments, and programs can specify their own definitions and expectations within these parameters. Second, a much wider array of instruments and techniques can be drawn upon for evaluation. We rely overwhelmingly on SETs in spite of the clear evidence in the literature that these tend to be biased, partial, and vague. It is vital that students be consulted; but this feedback also needs to be considered in relationship to other processes of evaluation that reflect faculty efforts and accounts of their objectives and methods in each course.

CONCLUSIONS

In 2019, the Academic Affairs committee of Faculty Assembly launched a review of how W&M evaluates teaching across all of the schools. To accomplish this, we combined two outward-facing approaches (literature review, case studies of recent reviews at other institutions), with two inward-facing approaches (survey of current W&M practices, review of current W&M policies).

Recommendations, which are listed in the Executive Summary, are wide-ranging and emphasize the importance of adopting a more “formative” approach to evaluation, which goes beyond simply evaluation and explicitly helps instructors improve as educators. This can lead to a more positive faculty growth mindset, one in which teaching ability is considered a work in progress. To accomplish this, we encourage each school to develop a broad, working definition of effective teaching and teaching expectations (as a function of career stage), and to communicate this widely to students, faculty, and administrators. We strongly recommend that W&M broadens the kinds of materials used to evaluating teaching BEYOND student course evaluation scores (which are fraught with potential bias). Although school and unit policies often refer to alternative approaches such as peer observation, feedback from administrators, self-assessment, teaching portfolios, assessment of course materials, and assessment of student work—these approaches are not always implemented *in practice*. When student evaluation scores ARE used, unidimensional ratings (such as Overall Teaching Effectiveness) should be down-weighted or eliminated in favor of questions based on established, multi-dimensional conceptualizations of teaching effectiveness, including Learning/value, Enthusiasm, Organization, Group interaction, Individual rapport, Breadth, Examinations/grading, Assignments, and Difficulty (Marsh, 1982). The distribution of student course evaluation scores, rather than mean scores, should be emphasized, and issues of sample size and random sampling must be explicitly addressed.

Some aspects of the overall teaching evaluation process should be determined centrally at W&M and adopted consistently across all schools, but the specific processes must be customized as needed to reflect the unique missions, goals, and student populations of different units. The process should, by definition, be developed by those who will be involved (i.e., including administrators, faculty, and students). We recommend that each school develop a committee to assess these recommendations and to determine how they can be implemented more effectively within the culture of that school. Ideally, we recommend that implementation take place AFTER COVID-related procedures have either been lifted or normalized. We view this analysis as a first step in strategically reassessing how W&M defines and assesses excellence in teaching.

APPENDICES

APPENDIX A: SURVEY OF PRACTICES-- SURVEY INSTRUMENT

Evaluating Teaching at William & Mary

Start of Block: Default Question Block

Q1

The following survey was developed by the Academic Affairs Committee of Faculty Assembly to collect data on how schools, departments, and programs at W&M evaluate teaching. Data from this survey will be combined with background from the higher education literature and case studies from other institutions, to help us identify most effective practices for teaching evaluation. The results will be disseminated as a white paper to the faculty, Faculty Assembly, Provost Agouris, and President Rowe. If you have questions on this work, please contact Rowan Lockwood (Chair of Academic Affairs, rxlock@wm.edu).

Q2 Which of the following **schools** do you represent? (Check one.)

- ☐ School of Business (1)
 - ☐ School of Education (2)
 - ☐ SMS (3)
 - ☐ Law School (4)
 - ☐ Arts & Sciences (5)
-

Display This Question:

If Which of the following schools do you represent? (Check one.) = Arts & Sciences

Or Which of the following schools do you represent? (Check one.) = SMS

Q3 Which **department or program** do you represent?

Display This Question:

If Which of the following schools do you represent? (Check one.) = Arts & Sciences

Or Which of the following schools do you represent? (Check one.) = SMS

Q4

Please take a moment to think about how your department or program (your unit) evaluates teaching on an ANNUAL BASIS. Your ratings should reflect how the system works in practice, not how you or your unit would like it to work. If you don't know the answer to a question, please feel free to say so.

Display This Question:

If Which of the following schools do you represent? (Check one.) != Arts & Sciences

Q5

Please take a moment to think about how your school (your unit) evaluates teaching on an ANNUAL BASIS. Your ratings should reflect how the system works in practice, not how you or your unit would like it to work. If you don't know the answer to a question, please feel free to say so.

Q6

For each category listed below, rate the extent to which this **information** is used when your unit evaluates teaching.

	Never Used (1)	Almost Never Used (2)	Occasionally Used (3)	Frequently Used (4)	Usually Used (5)	Almost Always Used (6)	Always Used (7)	Not Sure (8)
Student course evaluations (Q6_1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Peer classroom observation (i.e., by colleagues) (Q6_2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Self- evaluation (i.e., by oneself) (Q6_3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Assessment of course materials (i.e., syllabus, assignments) (Q6_4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Assessment of student work (Q6_5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student grade distributions (Q6_6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Quantity of teaching (e.g., course load, credit hours, contact hours, etc.) (Q6_7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please describe) (Q6_8)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q10

Beyond classroom performance, to what extent are each of these **other** activities included when your unit evaluates teaching?

	Never Included (1)	Almost Never Included (2)	Occasionally Included (3)	Frequently Included (4)	Usually Included (5)	Almost Always Included (6)	Always Included (7)	Not Sure (8)
Course development (i.e., individual courses) (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Curriculum development (i.e., new or revised programs, minors, majors) (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teaching innovation (i.e., new approaches, methods, technologies) (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Individual student support (i.e., office hours, mentoring) (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Promoting diversity and inclusion in teaching (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Course materials (6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please specify) (7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q7 For each of the **goals** listed below, rate the extent to which your unit's approach to evaluating teaching is effective in accomplishing it.

	Never Effective (1)	Almost Never Effective (3)	Occasionally Effective (4)	Frequently Effective (5)	Usually Effective (6)	Almost Always Effective (7)	Always Effective (8)	Not Applicable (9)
Faculty evaluation (e.g., annual merit reviews) (Q7_1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Faculty development (i.e., ongoing improvement in teaching ability and performance) (Q7_2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Curriculum development (i.e., ongoing improvement of courses and programs) (Q7_3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
External accreditation (Q7_4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (Please describe) (Q7_5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q8

Does your unit provide a description of **teaching expectations** (i.e., quality and/or quantity of teaching) for the purposes of evaluation? If so, please describe those expectations.

Q9 How are these teaching expectations communicated to faculty?

Q11

Think about how your unit uses **STUDENT COURSE EVALUATIONS** to evaluate teaching on an **ANNUAL BASIS**.

Q12

For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching.

	Never Used (1)	Almost Never Used (2)	Occasionally Used (3)	Frequently Used (4)	Usually Used (5)	Almost Always Used (6)	Always Used (7)	Not Sure (8)
Student evaluation scores (Q12_1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student evaluation comments (Q12_2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Other (please describe) (Q12_3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q13

Are some questions on student evaluation forms considered more important to your unit than others? If so, which ones?

Q14

For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching.

	Never Used (1)	Almost Never Used (2)	Occasionally Used (3)	Frequently Used (4)	Usually Used (5)	Almost Always Used (6)	Always Used (7)	Not Sure (8)
Mean student evaluation scores (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Median student evaluation scores (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Standard deviation of student evaluation scores (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Response rate of student evaluations (%) (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Course size (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Q15

For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching.

	Never Used (1)	Almost Never Used (2)	Occasionally Used (3)	Frequently Used (4)	Usually Used (5)	Almost Always Used (6)	Always Used (7)	Not Sure (8)
How student evaluation scores change over time (1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How student evaluation scores differ by course type (2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How student evaluation scores differ by course size (3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How student evaluation scores are affected by faculty gender (4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How student evaluation scores are affected by faculty race or ethnicity (5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How
student
evaluation
scores are
affected by
faculty
sexual
orientation,
gender
identity, or
expression
(6)

☐☐☐☐☐☐☐☐

How
student
evaluation
scores are
affected by
expected
grade (7)

☐☐☐☐☐☐☐☐

Q16

Does your unit set a **target minimum response rate** per course for student course evaluations? If so, what is it?

Q17 Think about how your unit **RESPONDS** to annual teaching evaluations.

Q18

Do faculty in your unit receive **feedback** from their annual evaluation related to their teaching? If so, please describe how they receive feedback.

Q19 Does your unit provide **mentoring** with respect to teaching? If so, please describe how mentoring is provided.

Q20

Think about how your unit evaluates teaching for DIFFERENT FACULTY AND CAREER STAGES.

Q21

Does your unit apply additional or different approaches from what you described above when evaluating teaching for **tenure and promotion**? If yes, please explain how they differ.

Q22

Does your unit apply additional or different approaches from what you described above when evaluating teaching for **tenure-eligible versus non-tenure-eligible** faculty? If yes, please explain how they differ.

Q23

Does your unit apply additional or different approaches from what you described above when evaluating teaching for **pre- versus post-tenure** faculty? If yes, please explain how they differ.

Q24 How does your unit **recognize or reward** improvement in teaching?

Q25 Please add any comments you wish to add regarding the evaluation of teaching in your unit.

End of Block: Default Question Block

APPENDIX B: SURVEY OF PRACTICES-- QUANTITATIVE DATA SUMMARY

Data in the following tables are summarized as %Always or Almost Always (%A), combined across all five schools, and separately for each school. Note that, for SOB and SOE, the evaluation of teaching does not vary across different units, and all survey questions were answered by a single Associate Dean for each school. For the SOL, two Associate Deans completed the survey and differed somewhat in their assessment of how the SOL approaches the evaluation of teaching. At SMS, the evaluation of teaching varies across the four academic departments and each department chair submitted survey data. Finally, for A&S, all chairs and program directors were asked to complete the survey. We received responses for 26 chairs and program directors, representing 19 departments or programs. Note that seven chairs or program directors chose not to share their unit name in the survey.

Q6 - For each category listed below, rate the extent to which this information is used when your unit evaluates teaching.

	All		A&S		SOB		SOE		SOL		SMS	
Question	n	%A	n	%A	n	%A	n	%A	n	%A	n	%A
Student course evaluations	34	94.1	26	92.3	1	100	1	100	2	100	4	100
Quantity of teaching (e.g., course load, credit hours, contact hours, etc.)	34	67.6	26	61.5	1	0	1	100	2	50	4	100
Assessment of course materials (i.e., syllabus, assignments)	34	35.3	26	34.6	1	100	1	100	2	50	4	0
Self-evaluation (i.e., by oneself)	34	29.4	26	23.1	1	0	1	0	2	50	4	75
Other (please describe)	9	26.5	9	100	0	0	0	0	0	0	0	0
Peer classroom observation (i.e., by colleagues)	34	14.7	26	19.2	1	0	1	0	2	0	4	0
Student grade distributions	34	14.7	26	15.4	1	0	1	0	2	0	4	0
Assessment of student work	34	8.8	26	7.7	1	0	1	0	2	0	4	0

Q10 - Beyond classroom performance, to what extent are each of these other activities included when your unit evaluates teaching?

	All		A&S		SOB		SOE		SOL		SMS	
Question	n	%A	n	%A	n	%A	n	%A	n	%A	n	%A

Course development (i.e., individual courses)	34	70.6	26	69.2	1	100	1	100	2	50	4	75
Curriculum development (i.e., new or revised programs, minors, majors)	34	47.1	26	46.2	1	100	1	0	2	0	4	75
Individual student support (i.e., office hours, mentoring)	33	45.5	26	46.2	1	0	0	0	2	50	4	50
Teaching innovation (i.e., new approaches, methods, technologies)	34	44.1	26	46.2	1	100	1	0	2	50	4	25
Promoting diversity and inclusion in teaching	34	14.7	26	11.5	1	0	1	0	2	50	4	25
Course materials	33	12.1	25	12.0	1	100	1	0	2	0	4	0
Other (please specify)	2	100	2	100	0	0	0	0	0	0	0	0

Q7 - For each of the goals listed below, rate the extent to which your unit's approach to evaluating teaching is effective in accomplishing it.

	All		A&S		SOB		SOE		SOL		SMS	
Question	n	%A	n	%A	n	%A	n	%A	n	%A	n	%A
Faculty evaluation (e.g., annual merit reviews)	32	34.4	24	20.8	1	100	1	100	2	0	4	100
External accreditation	18	22.2	16	18.8	1	0	1	0	2	0	0	0
Faculty development (i.e., ongoing improvement in teaching ability and performance)	29	10.7	20	10.0	1	0	1	0	1	0	4	0
Curriculum development (i.e., ongoing improvement of courses and programs)	29	7.1	21	4.8	1	100	1	0	0	0	4	0
Other (Please describe)	0	0	0	0	0	0	0	0	0	0	0	0

Q12 - For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching.

	All		A&S		SOB		SOE		SOL		SMS	
Question	n	%A	n	%A	n	%A	n	%A	n	%A	n	%A
Student evaluation scores	34	85.3	26	84.6	1	100	1	100	2	50	4	100

Student evaluation comments	34	61.8	26	61.5	1	0	1	100	2	50	4	75
Other (please describe)	3	0	3	0	0	0	1	0	0	0	0	0

Q14 - For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching.

	All		A&S		SOB		SOE		SOL		SMS	
Question	n	%A	n	%A	n	%A	n	%A	n	%A	n	%A
Mean student evaluation scores	33	84.8	26	88.5	1	100	1	0	2	50	3	100
Course size	28	53.6	22	45.5	1	100	1	0	1	100	3	100
Response rate of student evaluations (%)	33	39.4	26	38.5	1	0	1	0	2	50	3	66.7
Median student evaluation scores	32	31.3	25	28.0	1	0	1	0	2	50	3	66.7
Standard deviation of student evaluation scores	33	18.2	26	15.4	1	0	1	0	2	0	3	66.7

Q15 - For each of the categories listed below, rate the extent to which they are used when your unit evaluates teaching.

	All		A&S		SOB		SOE		SOL		SMS	
Question	n	%A	n	%A	n	%A	n	%A	n	%A	n	%A
How student evaluation scores differ by course size	34	47.1	26	46.2	1	100	1	0	2	50	4	50
How student evaluation scores differ by course type	34	44.1	26	46.2	1	0	1	0	2	50	4	50
How student evaluation scores change over time	34	26.5	26	23.1	1	0	1	100	2	50	4	25
How student evaluation scores are affected by faculty gender	34	8.8	26	3.8	1	0	1	0	2	50	4	25
How student evaluation scores are affected by faculty race or ethnicity	34	8.8	26	3.8	1	0	1	0	2	50	4	25
How student evaluation scores are affected by expected grade	32	3.1	26	3.8	1	0	1	0	1	0	3	0

How student evaluation scores are affected by faculty sexual orientation, gender identity, or expression	33	3.0	26	3.8	1	0	1	0	2	0	3	0
--	----	-----	----	-----	---	---	---	---	---	---	---	---

APPENDIX C: SURVEY OF PRACTICES-- QUALITATIVE RAW DATA

Question 8- Does your unit provide a description of teaching expectations (i.e., quality and/or quantity of teaching) for the purposes of evaluation? If so, please describe those expectations.

(29 respondents indicated concrete teaching expectations are communicated)

- **Core themes**
 - **Teaching loads**
 - **Annual review metrics**
 - **Teaching evaluations**
 - **Policies and Procedures/Handbook**
- Classical studies-adequate preparation, regular class meeting, regular evaluation of student work, explain grading system and expectations, organized and articulate lectures
- Sociology-From our procedures: 1. Evaluation of Teaching Quality Paralleling the variety of activities that faculty may employ in the process of teaching, the Department of Sociology recognizes that effective instruction is best evaluated on the basis of multiple sources of evidence..."teaching evaluations" ... Qualitative Evidence of Effective Teaching: In addition to the quantitative SETs, the Department of Sociology views several other factors as critical evidence for evaluating effective instruction..."student written comments"..."syllabi".."number of students"..."Difficulty of courses"..."Development of new courses or major transformations"..."interdisciplinary teaching"..."supervision and mentoring".."Classroom visitations"
- **Annual review**
 - SOE-2-2 load...Annual review metrics
 - SOB-We outline the process in our annual review letters, so everyone knows what matters and how much it counts.
- Economics- Tangible evidence of teaching, student evaluations...display: innovativeness...currentness...availability...appropriateness and supervision of independent study, honors theses, and departmental forums...conferences on teaching
- MDLL-Teaching evaluation.... Additional points awarded based on chair's evaluation of faculty performance in the following areas: new courses, honors thesis director, independent study courses, unpaid additional course
- A&S- Courses Taught (by semester, with cross-lists), Senior and honors theses advised, Service on honors committees, Service on Master's or doctoral degree committees at W&M or at other institutions, Independent study courses supervised, Number of concentration advisees, Number of pre-major advisees, Advising for Monroe Scholars or other student research projects, New courses taught or significantly revised, Overload teaching, Teaching/pedagogy programs participated in (at W&M or elsewhere), Guest lectures for W&M colleagues, Ensemble tours (domestic or international) planned and led by the faculty member for which venues were hired or solicited, Unusual or large-scale events (such as major works with orchestra) occurring on-campus which were planned and/or conducted by the faculty member, Awards for Teaching
- Sociology- "handbook of departmental procedures that guides how the PC evaluates teaching"...discount quantitative assessment because we know they are biased...merit teaching distribution comes from indicators that are less prone to such bias"

- English-specific criteria- preparation and regular meeting of classes; conscientious preparation of syllabi, reading lists, tests, and final examinations; and regular availability and helpfulness to students. Engaging in such activities as giving guest lectures or teaching colleagues' classes; service on examination committees; supervision of theses or independent study courses; creation and teaching of new courses; teaching a substantial number of students in classes with essay writing and other significant written work; taking students abroad or into the field; teaching summer school; teaching study abroad courses during the summer; teaching an overload during the regular year.
- SOL- Faculty Members' Responsibilities
 - American Bar Association Standard 404 requires that the full-time faculty of the SOL, as a collective body, shall fulfill the following core responsibilities as employees of the SOL:
 - Teaching, preparing for classes, being available for student consultation about those classes, assessing student performance in those classes, and remaining current in the subjects being taught...submit a teaching report
- History- We have teaching expectations in our Policies & Procedures geared to each level of class, from 100 to 400. This stipulates a range of reading and writing required at each level. We recently updated these expectations to reflect new kinds of student performance, including podcasts, vlogs, etc etc.
- Linguistics- Linguistics follows the English merit system for appointees affiliated with the English dept (which is most of them). The average of the courses' course means and instructor means is calculated and needs to be at least 4.5 to receive highest recognition. In addition, faculty must have contributed in other ways to get full credit in the teaching area-- for example, advised an honors study, taught 2-3credit independent studies, been on 8 thesis committees, etc.
(4 vague)
- Biology-“Nothing is written. It is negotiated with the chair”
- SMS has a faculty base expectations document that specifies quantity of teaching and advising. We do not have a document that discusses expectations for teaching quality.
- SMS-These have all been fairly long-standing agreements between a chair and a faculty member. There are rarely written records of these
- SMS-faculty are required to teach one (3-credit) graduate course per year... There is no stated expectation of the quality of teaching.
- SOL –annual credit hour requirement
- Art & Art History believe we handle this more verbally. We tend to accept a wide range of outcomes based on which material/area is teaching (outcomes tend to look different for advanced painting than for introductory 3D design, for example). But I think we're good about verbalizing concerns or issues with individual faculty members if/when they arise, and I feel we do so in a very collegial way.
- SMS -we are primarily a research unit, not a teaching unit. Our explicit expectations are based on hours taught per year
- Theatre- We rely heavily on student evaluation scores.
- Physics- 1 lecture course... as well as the regular the supervision of undergraduate and Ph.D. research student
(No- exemplar)

- Music and AMES- NO but we should. There are many things that are agreed upon but not written into policy and this creates bad feelings, misunderstandings, and faculty abuse. We also have a faculty member that generally teaches VERY few students and does almost no senior thesis/honors advising - and others who do a lot. So it's uneven and expectations might or might not change that.

Q9 How are these teaching expectation communicated to faculty?

- **Core Themes**
 - **Annual review**
 - **Policies and Procedures/handbook**
 - **Associate Dean/chair**
- Classical Studies-Departmental Policies on teaching effectiveness
- Bio-Only to individual faculty, not to the whole department. This is done in conversation. There is seldom a written record. ; By conversation from the chair to a faculty member
- SOE- The annual review form contains expectations regarding teaching expectations. Our school by-laws stipulate the expected teaching load.
- SMS-In addition to providing their own feedback, chairs communicate any recommendations I have back to their faculty as part of the evaluation process.
- SMS- Through the Policy & Procedures Handbook and at the annual performance evaluation. We discuss this in terms of the Planning Letter for the next calendar year and we also discuss this at Departmental Retreats, when we develop a 5-year teaching plan for the department faculty.
- SMS- The SMS policy and procedures document states this. Also, in evaluations each year, faculty teaching effort and course evaluations are communicated to them as a ranking relative to the distribution across the rest of the department
- SMS- teaching expectations are largely decided and communicated through the Associate Dean of Academic Studies, after consultation with Dept Chairs. Also at annual evaluation time by the Dept Chair.
- Economics-written doc, evaluation guidelines
- MDLL- they are spelled out in our manual and the department's best practices documents; they are also discussed regularly at department meetings. All incoming faculty are mentored by senior faculty: the expectation is that they will visit each other's classes in the first semester of appointment and receive formal feedback/exchange reflections from those class observations. Grade inflation is also addressed at department meetings and in retention and promotion reviews.
- SOL-schools handbook
- Sociology-they have access to the merit guidelines in our handbook of dpt procedures
- Music and AMES- through the chair; mostly through dept culture and meetings
- English-dept handbook
- Art & Art History- NTEs will almost always hear directly from the chair or associate chair. This is pretty true for TE faculty as well, although if there is greater overlap between 2 faculty members' area, it might make more sense for those 2 faculty to talk
- Art & Art History-They are spelled out in our dept procedures distributed to all faculty
- MLL-mll faculty handbook
- Theatre-we are working on a system that inform [faculty]

- Psychological Sciences-the chair works with faculty members regarding teaching load each year
- Physics-word of month
- A&S- job interview process and on the annual merit review form
- SOL- In the faculty handbook which is online and emailed to faculty annually. They also complete an annual report for the Dean, which is discussed in his annual review meetings that he holds with every full-time faculty member. Promotion and tenure standards also address teaching expectations with the aspiration either excellence in teaching or making progress towards excellence in teaching.
- Linguistics- Through how the points are calculated for the teaching section of the merit evaluation.
- History- They are in our Policies & Procedures, and new faculty receive this section from the Director of Undergraduate Studies when they are putting their courses together.

They are not:

- Sociology-They receive their merit scores but there is no formal process for discussions about teaching to take place

Q13 Are some questions on student evaluations forms considered more important to your unit than others? If so, which ones?

- **Core Questions**
 - overall quality of the course
 - the overall quality of the instructor
 - Preparedness
 - Effectiveness
 - Educational value
- Sociology-All of these items may be considered in evaluating a faculty member for merit. However, at a bare minimum, the two summary questions that ask students to assess the overall quality of the course and the overall quality of the instructor must be utilized.
- Sociology- We look at the overall course and instructor questions.
- Biology- Yes, we pay attention to three questions in particular, "teaching effectiveness", "how much did you learn", and "overall evaluation of the course"
- Philosophy- For the numerical part, we focus *almost* exclusively on the overall evaluation of course and instructor.
- Anthropology- We consider the scores holistically, but the 'overall teaching effectiveness score' receives the most attention.
- SOE- all questions
- SMS- I consider some to be more important, but I don't have time to go into the details right now.
- SMS- Questions concerning the preparedness of the faculty (well organized lectures, etc.).
- SMS- Yes, the score that is used in evaluations in my department is the "overall evaluation of instructor".
- SMS- The "overall" /summary questions for instructor and course.
- SOB- We always use two questions (Instructor Effectiveness and Overall Educational Value) and we have in the past used a third (Instructor's Attitude Toward Students).

- Biology- We pay attention to three questions in particular: 'instructor performance', 'how much did you learn?', and 'overall course evaluation'.
- Economics-Overall course and overall instructor scores
- MDLL- The teaching effectiveness score carries a lot of weight since it is the sole number requested by the administration for personnel reviews.
- SOL- Overall effectiveness; course was intellectually demanding.
- A&S-teaching effectiveness
- A&S- We look most closely at overall teaching and overall course
- A&S- Yes question #11 Overall Effectiveness of the Instructor
- A&S-Overall teaching effectiveness and how would you rate this course overall
- A&S-It really comes down to the last two questions
- Philosophy- Overall course and overall instructor are typically by far the most important.
- English- How would you rate this instructor's overall teaching effectiveness?
How would you rate this course overall?
- Mathematics- How would you rate this instructor's overall teaching effectiveness?
- Art & Art History- We look most closely at overall teaching and overall course
- Psychological Sciences- the questions about the instructor and overall course are included in mid-tenure, tenure, and promotion cases.
- Art & Art History- Information on availability to students and returning assignments in a timely manner; the organization and delivery of class material
- Physics- We look at the evaluations holistically. The overall teaching effectiveness score is useful -- there has always been a good correlation between that number and other metrics available to us. The faculty who complain the most frequently and vociferously about the weight placed on the overall teaching effectiveness score have, in my experience, also been the worst instructors in my department.
- SOL- Narrative comments to open questions are usually the most helpful.
- Linguistics- Overall course and instructor means (last two categories on the evaluation forms)
- History- Not really. We try to look at them holistically and take account of both positive and negative comments and scores.

Q16- Does your unit set a target minimum response rate per course for SETs? If so, what is it?

NO

- Classical studies- not but we should
- Sociology-no
- Sociology-no
- Biology-no
- Philosophy-no
- Philosophy-no
- Anthropology-no though we probably should
- SOE-no identified minimum
- SMS-we hold back the evals if there are 3 or fewer students in the course
- SMS-no, we do not have target minimum response rate
- SMS-no
- SMS-no some of our classes at SMS are very small

- SOB- 5
- Art & Art History – no. there;s too much variation
- Art & Art History- No, but we take it into consideration I don't think we generate forms for class sizes of 2 or under (which is common for "stacked" courses)
- Art& Art History – we have no set minimum response rate
- SOL- We take steps to get high rates, like setting aside class time, but there is no particular number.
- Music & AMES- no
- A & S-no
- English- We used to say 75% but we decided to stop doign that a few years ago when response rates became wildy uneven after we switched to electronic evaluations.
- A & S-no
- A&S- we have no minimum but treat low response rates with caution
- Mathematics-no
- Theatre-no
- Psychological sciences- our goal is a 70% response rate
- Physics-no
- SOL- No, but we do give class time to do online evaluations in order to maximize response rate.
- Linguistics-no
- History-no

Q18- Do faculty in your unit receive feedback from their annual evaluation related to their teaching? If so, please describe how they receive feedback.

- **Core themes**
 - **Merit score**
 - **Not formally**
 - **Annual meetings**
- Classical studies- usually positive comments in the chair's merit response
- Sociology- Only in the form of a merit score. If there is some issue that the personnel committee feels should be addressed, they consult with the Chair and the Chair speaks with the faculty member.
- Biology- No, not formally. Though some chairs in the past have talked with some faculty members.
- Philosophy- If the responses are remarkably high or remarkably low, this is reflected in the paragraph of narrative evaluation. On some very rare occasions, the chair will speak with a faculty member to address specific concerns.
- Anthropology-no
- SOE- Sometimes discussed in annual meetings with faculty members
- SOB- They receive oral, written, and numeric feedback.
- Biology- No. They receive a merit score, and that's it.
- SMS- Yes. I prepare a summary of the number of courses delivered by the Department's faculty in the calendar year. I also prepare individual statements for each faculty member concerning the positive and negative feedback that I glean from student evaluations of their courses and from the Dean of Academic Studies. If applicable, I point out areas of improvement so that faculty can focus their efforts accordingly. All of this is received in

written form (memo) and also discussed verbally during the annual performance evaluation.

- SMS- Their written annual evaluations include a paragraph specifically evaluating their teaching. They also meet with the department chair to go over their evaluation.
- SMS- Yes, there is an in-person meeting to discuss teaching along with all the other aspects of the evaluation (research et.), and the Chair provides brief written comments summarizing the evaluation.
- Economics- Sometimes discussed in annual meetings with faculty members
- MDLL- NTE faculty with teaching evaluations below 3.8 average teaching effectiveness score are mentored by colleagues.
Junior faculty with low teaching evaluation scores are mentored by colleagues.
- SOL- The annual meeting with the Dean frequently includes discussion of teaching effectiveness, new courses, teaching innovations.
- A & S-as a merit score
- A&S- no
- A&S- Yes, especially younger ones; usually by face-to-face conversation with chair
- A & S- Yes. In addition to a numeric score, faculty receive a 1-3 paragraph explanation of the score.
- Sociology- We have departmental conversations about teaching, and pre-tenure faculty have mentors who discuss their performance with them (along with the Chair doing so). But this isn't part of annual evaluation (by which I assume you mean merit). We really look to moments of review or promotion, and not annual merit, as the time to check in on teaching effectiveness; and we rely much more on alternative indicators.
- Philosophy- The chair's narrative assessment provides such feedback. It is typically quite brief, but might highlight recurring themes in the written comments from students.
- Music and AMES- Depends on the chair! And whether the faculty member is new or up for evaluation.
- English- The Chair disseminates their merit score for teaching and if the score suggests there are major problems, the Chair meets with the faculty member.
- Art & Art History- Faculty receive a written response from the chair summarizing their merit scores. If there is a problem with a faculty members teaching, someone from Personnel, and probably the chair/associate chair will discuss it in person to avoid lack of clarity or a sense of hostility.
- Art & Art History- They receive numerical scores on their merit report that correspond to the excellent, good, satisfactory and unsatisfactory categories.

Feedback is offered to NTEs and tenure eligible faculty, and when a persistent problem needs to be corrected.

- Mathematics-no
- Theatre- We need to have such a process or system.
- Neuroscience-no
- Psychological Sciences- They received a score for teaching from our Personnel Committee.
- Physics- The only feedback they receive is the report showing numerical student evaluation scores as well as the student comments. They are provided their peer merit scores for teaching upon request as well as the departmental average.

- SOL- Yes, in an individual meeting with the Dean, and either the Vice Dean or the Associate Dean for Research & Faculty Development.
- Linguistics-no, just the numeric score
- History-In some cases yes, if there is a major issue. This will be between the Chair and the faculty member.

Q19- Does your unit provide mentoring with respect to teaching? If so, please describe how mentoring is provided.

- **Core themes:**
 - **Informally**
 - **Pre-tenure faculty/new faculty**
 - **As needed**
- Classical studies- informally, we observe classes of NTEs and new TEs, provide feedback
- Sociology- Yes. We have a meeting or two every semester called "Teaching Enhancement." This is a meeting where we only touch on teaching and student/mentoring issues. All NTE and TE faculty attend these meetings and they have been highly productive.
- Sociology- Yes, in two ways. Pre-tenure faculty are encouraged to have a senior colleague sit in on their classes and provide written feedback. Also, every semester we have a faculty meeting devoted to teaching enhancement (NTE faculty are included); these are organized around a theme and we discuss our particular challenges and strategies.
- Biology- Each faculty member is assigned a mentor. People are encouraged to help those who might be struggling.
- Biology- Yes. Every new faculty member is assigned a mentor. The mentor is meant to talk to their mentee about teaching. Whether that happens and the quality of the interactions are highly variable. But the intent is for everyone to have active mentoring.
- Philosophy-Before tenure, faculty classes are visited by tenured faculty as part of the run-up to tenure, and this can result in some feedback; presently these visits are primarily concentrated in the year prior to tenure. Junior faculty are encouraged to participate in the teaching project.
- Philosophy- On a case by case basis, and only when there seems to be a problem, the chair might approach a faculty member. Virtually always this will be a junior faculty member.
- Anthropology- Before tenure, faculty classes are visited by tenured faculty as part of the run-up to tenure, and this can result in some feedback; presently these visits are primarily concentrated in the year prior to tenure. Junior faculty are encouraged to participate in the teaching project.
- SOE-no
- SMS- most often accomplished through team teaching in our core courses
- SMS- In the past, some faculty have received mentoring from the Department Chair (before I became chair in 2018). The mentor would sit in on the class and provide verbal feedback to the instructor. I don't think this was very effective - I still see the same problems with the instructor who is not fully committed to teaching. I've been

contemplating other approaches, but the 'word is out' that the instructor is not very effective, so none of the graduate students want to sign up for his course.

- SMS- We commonly have faculty mentors paired with new faculty members. They mentor on all topics. But I don't think we have formal mentoring specifically for teaching.
- SMS- For new faculty in our larger Fundamentals of Marien Science courses, we pair them with experienced senior faculty so the first time they do the course it is co-taught. The rest is informal mentoring - usually it is left up to the mentee to seek this out.
- SOB- We have a teaching coach, and the AD of Faculty works with faculty to ID problem areas and to set a plan of recovery for faculty who are struggling in the classroom.
- Economics- not unless requested
- MDLL- peer observation; may seminars; co-teaching; syllabi development and review of course assignments; department workshops to address thorny and shared/common teaching issues.
- SOL- For untenured faculty, an assigned mentor and the Associate Dean visit classes, review materials, and provide advice. The third-year retention view produces a mentoring report that addresses teaching as well as research and service.
For tenured faculty, the administration intervenes when there seems to be a problem. We have faculty lunches devoted to teaching several times per year.
- A&S-only as needed
- A&S- For younger/beginning faculty or faculty in their probationary period
- A&S- Yes, if deemed necessary. Usually by working with senior member of faculty.
- A&S- Each new faculty member is assigned a senior faculty mentor. Mentor topics usually include teaching.
- Music and AMES- Theoretically yes and certainly among faculty on an informal basis but right now, not really. Again, depends on the chair. Right now, no!
- English- New faculty are mentored by the Chair and often by another mentor chosen either by them or by the Chair with their input. And see above for established faculty members who are struggling with their teaching.
- Art & Art History- We talk a lot about teaching in informal settings. We also do a lot of peer observation (at TE and NTE levels). We also invite colleagues into our classrooms to see effective teaching if a faculty member is struggling. We share resources such as syllabi, handouts, image collections, etc.
- Art & Art History- Mentoring is informal. We have no formalized mentorship. We also do periodic peer teaching observations for tenure eligible faculty and lecturers prior to promotion.
- Mathematics-no
- Theatre-not formally
- Neuroscience-no
- Psychological Sciences-we have a pre-tenure mentoring program when teaching would be discussed
- Physics- We have been providing new faculty with assigned mentors within the department. They can provide teaching advise, if it is requested.
- SOL- Junior faculty members on the tenure or security of position track are assigned a tenured mentor until they reach tenure/security of position.

- Linguistics- We have a lot of informal mentoring, the most formal version of which was regular meetings for circa 5 years of the "220 steering committee" made up of those faculty teaching the introductory course (LING220 Study of Language) where we shared resources and discussed concerns. This no longer meets because we developed a corequisite workshop for the course and so we meet weekly to prep for that and also everyone teaching the course has now done so many times and so there is not the same need for active discussion and mentoring. Junior faculty often bring questions/concerns about students to me and I am very happy to discuss them.
- History- Yes. History faculty often sit in on each other's classes, esp. of new faculty, and offer feedback. As Chair, I usually select especially innovative and committed teachers in the department who are especially good at this.

Q21- Does your unit apply additional or different approaches from what you described above when evaluating teaching for tenure and promotion? If yes, please explain how they differ.

- **Core themes**
 - **Evaluation of teaching materials**
 - **For tenure/pre-tenure**
 - **During merit review**
 - **Class observation**
- Classical Studies- no one observed any of classes at any stage: we observe classes when asked
- Biology- We have teaching visitations in the year before tenure
- Biology- Yes. We evaluate teaching materials (syllabus and assessments, mostly), and we have visitations to the faculty member's classroom by a couple of tenured faculty--most often those are people on our personnel committee and it happens in the year before tenure review.
- Philosophy- For tenure, we try to get all senior faculty to visit a range of courses. This provided more material for evaluation than SETs, but tenure letters also include discussion of those evaluation numbers. In tenure letters, there is more attention paid to trends and trajectories in those numbers, and to issues of class size, class topic, and so on.
- Philosophy- Pre-tenure, the senior faculty come to visit classes. Most of us use a standard form to record content, student behavior, teaching techniques, and so on. Usually all the faculty visit at least one class, but this also typically happens in the year prior to coming up for tenure, rather than from the very beginning of someone's employment at W&M.
- Anthropology- During our annual merit review the focus is on the types of courses taught and student evaluations. For tenure and promotion we incorporate a broader range of evidence including self assessment, peer observation, improvement, innovation, and the types of courses taught.
- SOE-no
- SOB- Outside the annual review process, our faculty Personnel Committee reviews faculty on a rolling basis. This review does not feed into merit, but it does feed into renewal, promotion, and tenure decisions. This process involves Teaching Review Committees for each faculty member being reviewed. These committees attend at least

one session and review faculty on a host of teaching metrics, including:

-clarity of course goals

-rigor

-pedagogy

-syllabus construction

-in-class effectiveness (e.g., presentation, response to student Qs, enthusiasm, pacing, command of material, use of examples, student interaction, student participation, and discussion).

- SMS- Generally, no, except that we recognize that first-time instructors are not likely to have the best student evaluations. We do expect those evaluations to improve over time. We also do not overload faculty with teaching requirements if he/she is coming up for tenure and promotion.
- SMS- In evaluation tenure/promotion, our department letter, which comes from the department's faculty as a whole, is more likely to emphasize strengths compared to the ranking approach emphasized in annual evaluations. Also, our Associate Dean of Academic Studies writes a separate letter that is important to the tenure/promotion process. The Associate Dean does not always weigh in annual evaluations, although Assoc Dean can contribute to the annual evaluation if s/he would like to.
- SMS- We like to see that a faculty member has designed a new course/s or has taken over significant responsibility for an existing course.
- Economics-For tenure and promotion, trends over time are also important.
- MDLL-more attention is paid to
 - 1) changes implemented by faculty under review to address teaching challenges;
 - 2) help sought through existing programs like teaching project
 - 3) improvement in teaching scores and comments for repeated courses
- SOL- Tenure and promotion reviews involve extensive review of teaching. Committee members and some members of the faculty at large visit classes (often more than one kind of class, e.g. if the candidate teaches both large classes and a seminar). The committee reviews teaching evaluations and solicits additional student comments. The third-year retention review produces a mentoring report for the candidate that addresses teaching (among other topics).
- A&S- Yes. Faculty will observed classroom teaching.
- A&S-no
- A&S-no
- A&S- Yes, we conduct two separate peer observations, conducted over two separate semesters.
- Sociology- As I just explained, we think of the run-up to tenure and promotion as an opportunity to give feedback and guidance in order to support colleagues in that process.
- Music and AMES- We make sure that faculty have peer reviews.
- English- Yes, very different. We read all student comments, look at all scores (ie answers to all questions) from all courses under review, read and assess all syllabi and selected assignments, and of course, pay attention to the faculty member's own evaluation of their teaching in the narrative statement.
- Art & Art History- When a TE goes up for tenure or promotion, we include their peer observations since their hire, or for promotion to full, since their tenure. We make sure a faculty member has a robust dossier of observations in the years leading up to promotion

or tenure. We tend to have every tenured faculty member observe every NTE or not-yet-tenured TE faculty at least once for their dossier. When I went up for tenure, some faculty observed me twice over those first 6 years. Promotion to full is a newer endeavor for us, as we had many retirements in the past 8-12 years. But I believe our practice is similar for promotion to full.

- Peer teaching observations are required for tenure and promotion but not for annual merit.
- Mathematics- Some years ago, we used senior survey for major students, as the students usually have taken classes with different professors. But that practice has stopped maybe 10 years ago.
- Theatre-no
- Psychological Sciences-we also include peer evaluations
- Physics- Tenure and promotion cases include the "second form of evaluation" which is usually an unannounced classroom visit. It is generally useless, since it is a single visit compared to 5+ years of SETs. However, it does satisfy the RPT requirements.
- SOL- Criteria for Retention, Promotion, and Tenure for Tenured and Tenure-Eligible Faculty

A. General Criteria: All decisions regarding retention, promotion, and tenure will be based

- on an evaluation of a candidate's achievements with respect to the following factors:
 1. Teaching Effectiveness: Retention, promotion, and tenure decisions will be based, in part, on an evaluation of a candidate's competence as a teacher, including teaching-related activities outside the classroom. Evaluation of teaching effectiveness will be based on both peer and student opinion. Evaluation of teaching effectiveness will involve class visitations by faculty members, review of course evaluations, and solicitation of written student opinion. The Committee shall establish class visitation guidelines designed to minimize classroom disruption.

Tenure

a. The award of permanent tenure signifies the attainment of academic excellence, and shall be based on the criteria specified in Section IV(A). A recommendation for the award of tenure requires a judgment that the candidate has achieved excellence in teaching or scholarship, and has achieved proficiency with respect to all other applicable criteria. A recommendation for the award of tenure also requires a judgment that the scholarly publications of the candidate constitute a significant contribution to the body of legal or other scholarly literature.

- Linguistics- Yes, for tenure (and promotion?) we also consider teaching evaluations comments, course syllabi, exams, and classroom observations.
- History-no

Q22-Does your unit apply additional or different approaches from what you described above when evaluating teaching for tenure-eligible versus non-tenure-eligible faculty? If yes, please explain how they differ.

- Core theme
 - No

- Classical studies- no, same metrics apply; good teaching is good teaching regardless of status
- Sociology-no
- Biology- No. We are new to having lecturers, but have completed teaching visitations too
- Biology- So far, we have paid more attention to evaluating the teaching of our NTE than our TE faculty. Each NTE faculty has received classroom visitations followed by feedback debriefs. This doesn't happen for TE faculty until they come up for tenure (see above).
- Philosophy-we use the same criteria
- Anthropology- No. We have only one lecturer, and she is evaluated the same way tenure-line faculty are.
- SOE-no
- SOB-no
- SMS- No. We make no distinction here.
- SMS- No. Our annual ranking and evaluation approach in our department is exactly the same for TTE vs. NTE (i.e., "Research") faculty. The ranking may be weighted differently for any faculty member (TTE or NTE) depending on what their profile is for % of effort in education.
- SMS- Not really applicable as our non-tenure eligible faculty rarely teach courses, although they are often involved in mentoring of graduate students.
- MDLL- please see above excerpts from our manual.
Lecturers are formally reviewed twice prior to promotion to senior lecturer: in these reviews, teaching performance is carefully evaluated, including grade distribution, teaching effectiveness scores, syllabi, curricular development
- **SOL- For NTE faculty, there is less involvement by the faculty at large; evaluation and oversight of teaching is more concentrated in the administration.**
- **A&S- Yes. NTEs only submit a yearly merit dossier, including teaching evaluations. They are not evaluated by peers, only the chair.**
- A&S-no
- A&S-no
- **A&S-yes. Different set of criteria for NTEs**
- **A&S- Yes. Non-tenure eligible faculty are assessed in the following manner:**
 - Exceeds Expectations. This ranking indicates an excellent performance that far exceeds departmental expectations in all essential areas of responsibility and in contributions to overall departmental curricular goals. Performance must be judged as exceeds expectation through peer teaching observations and SETs.
These quantifiable achievements must include evidence of:
 1. Extraordinary contributions to the overall program curriculum and overall teaching goals of the department
 2. Student evaluations (for course and instructor) that are significantly higher than similar courses and confirmed by course evaluation comments
 3. Clear, rigorous, and engaging syllabi along with samples of student work or tests

4. Outstanding peer teaching reviews that confirm a high level of classroom performance, academic rigor, and delivery of syllabi and agreed upon shared educational values for the department
 5. Evidence of strong active teaching engagement outside the classroom which could include independent studies, honors theses, participation in university teaching workshops, participation and significant critical contributions in departmental student reviews, writing recommendation letters for students for university honors and graduate study, and attendance for student exhibitions and presentations.
- Meets Expectations. This ranking indicates clear progress towards excellence and/or good performance that consistently meets expectations, and at times possibly exceeds expectations, concerning contributions to overall departmental curricular goals. Performance is judged as meets expectation through peer teaching observations and SETs.

These quantifiable achievements must include evidence of:

1. Good contributions to the overall program curriculum and overall teaching goals of the department
 2. Student evaluations (for course and instructor) that meet the department mean for similar courses and confirmed by course evaluation comments
 3. clear, rigorous, and engaging syllabi along with samples of student work or tests
 4. Good peer teaching reviews that confirm a high level of classroom performance, academic rigor, and delivery of syllabi and agreed upon shared educational values for the department
 5. Evidence of active teaching engagement outside the classroom which could include independent studies, participation in university teaching workshops, participation and critical contributions in departmental student reviews, writing recommendation letters for students for on university honors and graduate study, and attendance for student exhibitions and presentations.
- Fails to Meet Expectations. This ranking indicates lack of reasonable progress toward meeting the critical goals of the department and/or unsatisfactory performance, judged on the basis of minimum standards and expectations of the department and contributions to the overall departmental curricular goals. Performance must be judged as did not meet expectations if continued performance trajectory is unlikely to result favorably for promotion to senior lecturer. This ranking would be determined by:
1. Uneven contributions to the overall program curriculum and overall teaching goals of the department
 2. Student evaluations (for course and instructor) that fall significantly below the department mean for similar courses and confirmed by course evaluation comments
 3. Problematic syllabi along with weak examples of student work or tests.
 4. Poor peer teaching reviews in terms of classroom performance, academic rigor, and delivery of syllabi and agreed upon shared educational values for the department.

5. Lack of evidence of active teaching engagement outside the classroom which could include lack of or uneven participation and critical contributions in departmental student reviews and poor attendance for student exhibitions and presentations.
- Sociology- We apply the same procedures for TTE and NTE in annual merit reviews and in the teaching enhancement faculty meetings.
 - Philosophy-The above approach only applies to TE faculty.
 - Music and AMES-not sure
 - English- Yes, the approaches I described above are for tenure-eligible and tenured faculty. To be eligible for reappointment, NTE faculty must demonstrate a record of effective teaching as reflected in course evaluations and in such other forms of evaluation as are generally employed by the Department. Numerical student evaluation scores will be a primary factor of evaluation, but other teaching materials may be consulted by the Personnel Committee, including syllabi, tests, and graded assignments, and reports of classroom visitations. If an adjunct, lecturer, or senior lecturer's teaching fails to meet expectations for any one year, the Department Chair will meet with the person to discuss the situation and to set up a plan for rectifying it in conjunction with the Committee on the Evaluation and Improvement of Teaching. Two out of three unsatisfactory reviews will normally lead to an appointment's not being renewed.
The annual merit evaluation of NTE faculty will come to one of three conclusions: Exceeds Expectations; Meets Expectations; Fails to Meet Expectations.
 - Art & Art History- The only major difference I can think of is that our last NTE promotion (our first since the initial wave when senior lecturer was created) had observations from not-yet-tenured TE faculty, as well as from tenured faculty. Not-yet-tenured faculty have not historically evaluated through observation other faculty members going up for tenure.
 - Mathematics-no
 - Theatre-no
 - Psychological Sciences- The mean student evaluation scores are translated into a 1, 2, or 3 for NTE faculty.
 - Physics- We have only recently hired our first NTE Lecturer, so the procedures for evaluating performance are being developed now.
 - SOL- Teaching is evaluated the same from what is described; for security of position, teaching is evaluated the same as those on the tenure track.
 - Linguistics- NTE merit guidelines for Exceeds Expectations:
Exceeds Expectations: Mean scores on student evaluations of at least 3.9 on the questions relating to overall teaching effectiveness and overall quality of course; and either effective service on a departmental standing or ad hoc committee, or advising of at least one Honors thesis, independent study, internship, Monroe project, or other student undergraduate research; other teaching material as described above may be considered in the Personnel Committee's evaluation and will be incorporated into its overall conclusion.
 - History- No, the same procedures apply, although we are more proactive observing their classes early on and following up with them if there are any issues.

Q23- Does your unit apply additional or different approaches from what you described above when evaluating teaching for pre- versus post-tenure faculty? If yes, please explain how they differ.

- **Core Themes**
 - **Classroom visitations**
 - **Focus on improvement**
 - **No (approx.. half)**
- Classical Studies- no difference
- Sociology-no
- Sociology- No, I don't think we do. (Though I believe it's less often the case that post-tenure faculty invite other faculty into their classrooms for observation.)
- Biology- More attention is given to pre-tenure faculty. There have not been routine teaching visitations for post-tenure faculty.
- Philosophy- We use the same criteria, though we do not have class visits for tenured faculty. Nor to we have class visits for senior lecturers.
- Philosophy- Post-tenure there are typically no class visits from peers until a professor wants to come up for Full, in which case the same sort of procedure is used, though fewer faculty typically participate.
- Anthropology- For pre-tenure faculty there's a greater focus on improvement and innovation.
- SOE-no
- SOB-no
- Biology- Post-tenure faculty do not receive classroom visitations and are evaluated only by the student evaluation forms. When someone comes up for promotion from associate to full professor we will evaluate course materials again, but it is rare for their to be classroom visitations. The biology faculty have discussed this recently and voted against the regularization of classroom visitations.
- SMS- No, other than to note that pre-tenure faculty are not likely to receive the best student evaluations, particularly for a course that was taught for the first time.
- SMS-no
- SMS- More detailed feedback, attention is given to pre-tenure faculty.
- Mathematics-no
- MDLL- post-tenure faculty are not reviewed except for merit yearly and when they come up for promotion to full.
- SOL- Yes. As noted above, junior faculty have a mentor, regular class visits, and regular meetings with the mentor and Associate Dean.
- A&S-yes, same as above
- A&S- No, just little more forgiving of junior faculty members
- A&S- Yes. There is no additional information considered above and beyond teaching evaluations and rate of response in Post-Tenure Review.
- Theatre-no
- Psychological sciences- no
- Music&AMES-not really
- English- We use the same approaches for pre-tenure and post-tenure faculty.
- Art & Art History- Not through merit. We do not have very clear procedures for promotion to full. We had our first 2 cases in decades (I believe 4 and 5 years ago). But I

think our understanding is that all tenured faculty members will observe and evaluate faculty as they go up for full. But that only other full professors will vote (or evaluate) the teaching in promotion-to-full cases.

- Art & Art History-no
- Physics-no
- SOL- Promotion to the rank of Associate Professor signifies the attainment of academic proficiency and shall be based on the criteria specified in Section IV(A). A recommendation for promotion to the rank of Associate Professor requires a judgment that the candidate has achieved proficiency as a teacher, a scholar, and with respect to all other criteria for retention, promotion and tenure, and has made satisfactory progress toward tenure, demonstrating potential for excellence as a scholar or a teacher.
- Linguistics-no
- History-no, same procedures

Q24-How does your unit recognize or reward improvement in teaching?

- **Core Themes**
 - **Higher merit scores (raises)**
 - **University nominations for teaching awards**
- Classical Studies- higher merit scores (for what that's worth)
- Sociology-yearly merit
- Sociology- One way is, at the pre-tenure level, to have a faculty observer visit a class more than once to be able to write a narrative of improvement in teaching (where applicable). I'm not sure we really reward improvement in other ways, because we have only the biased student assessments. In practice we seem to really reward effort in non-classroom teaching (overseeing student research, internships, etc), and in course innovation/design.
- Biology- Much of tenure and promotion discussion is based on demonstration of improvement of teaching. We value that highly.
- Biology- It is both recognized and rewarded through merit evaluation and at times of promotion. There is no internal teaching award or prize for our department. But we do regularly nominate people for A&S and university teaching prizes, so that is some form of reward and recognition for improving teaching.
- Philosophy- I don't think we notice trends unless they are stark, which means there was initially some problem, even if a slight one. In that sort of case, a praising remark in an annual evaluation is probably the only recognition.
- Philosophy- The chair's comments on the annual evaluation can sometimes do this. There are no explicit rewards, though.
- Anthropology- With higher merit score and nomination for university-wide teaching awards.
- SOE- Factors into raises and promotions when relevant.
- SOB- Higher teaching scores, calculated from all the dimensions above, lead to better merit scores, which leads to larger raises and more opportunities to teach overload.
- SMS- Good question! I really do not know, other than to note it in the annual performance evaluation.
- SMS- For outstanding faculty, we nominate them for SMS and University teaching awards and other SMS and University awards that consider strength of teaching. For

systematic improvement in teaching, we may also weight more recent contributions more heavily. (Our annual evaluation is usually a 4-year running average.)

- SMS- Mostly just with improved annual evaluations in that category.
- Economics-merit scores
- MDLL-we don't
- SOL- Improvement in teaching is viewed favorably by the faculty in tenure and promotion reviews. Recent strong performance will tend to overcome early rockiness. We have several teaching awards, some chosen by students and some by faculty or administration. Those recognize excellence and innovation.
- A&S- only as a better teaching merit score
- A&S- Merit evaluation, mention in Global Voices newsletter, or recognition at Faculty meetings.
- A&S-merit raises
- A&S-only in the merit review process
- Music and AMES- depends on the chair
- English- Hmm. Higher merit score? Approving remarks from the Chair?
- Art & Art History- It really depends. The big example I can think of was an NTE who consistently had good outcomes, but students complained about lack of clarity in expectations. That faculty member attended teaching workshops, was mentored individually by tenured faculty, and was invited to observe other faculty in the classroom. This led to great improvements in this individual's teaching, and they have since seen improved merit scores.
- Art & Art History-not consistently
- Mathematics- We have an annual departmental teaching award which is determined by student nominations and a committee consisting of students, one faculty, department chair and representative of W&M Foundation (the award money is from a private donation).
- Psychological Sciences-this factor is examined in tenure cases
- Physics- It is rewarded with higher merit scores. Some day, those scores may even correlate with pay raises.
- SOL- There are several ways - two teaching awards, the McGlothlin Teaching Award and the Kelly Teaching Award. Both involve extra monetary compensation and the latter involves presentation of two or three workshops for the rest of the faculty on best practices in teaching. Chaired professorships are also awarded to excellent teachers and scholars.
- Linguistics- It's only considered as something that should be part of a tenure package.
- History- Not well enough, really. We will put forward excellent teachers for awards from year to year, but this is a low-yield option. The Department will support faculty who want to undertake extra training where available and if a cost is involved.

APPENDIX D: CASE STUDIES-- INTERVIEW QUESTIONS

Review-based questions

- What prompted the review?
- Who was involved in the review?
- What was the timeline of the review?
- What were the goals/objectives of the review?
- What were the most useful resources (on or off campus) identified by the reviewers?
- What data were collected by the reviewers?
- What were the strengths of the former system for evaluating teaching? The weaknesses?
- What were the major recommendations of the review?
- How were these recommendations communicated to administrators, faculty, and students?
- What do you wish you had done differently with respect to the review?
- What do you wish you knew at the beginning of the review-- that you know now?

Implementation-based questions

- To what extent were the recommendations implemented?
- Who was involved in the implementation?
- What was the timeline of the implementation?
- What were the goals/objectives of the implementation?
- How was the implementation assessed?
- What were the strengths of the new system for evaluating teaching? The weaknesses?
- What do you wish you had done differently with respect to implementation?
- What do you wish you knew at the beginning of the implementation-- that you know now?
- How did you make decisions about allocating resources to both develop and, especially, to implement a new system ongoing?

APPENDIX E: CASE STUDIES—RAW DATA

Smith College Interview

Date: February 6, 2020

Time: 10-10:45am

Tool: Zoom video and audio

Interviewer: Rowan Lockwood

Interviewee: Floyd Cheung (previously head of Center for Teaching and Learning, now VP for Equity and Inclusion, Professor of English Language & Literature and American Studies)

Supplementary Info: See Google Drive appendix for copy of Final Report of Ad Hoc Review Committee, New Student Feedback Form, Old Course Evaluation Form, and Example Feedback.

Review-based questions

1. What prompted the review?

Course evaluations used to be administrated by the student side of the house. The focus was very much on student, as opposed to faculty development. When Katherine Rowe was hired as Provost (2014), she pushed for Smith to take a closer look at how they evaluate teaching, and charged them with re-evaluating that approach every 2-3 years.

2. Who was involved in the review?

Floyd was chair of the ad hoc review committee. The committee included faculty and students, with a particular focus on people who know a lot about teaching and learning.

3. What was the timeline of the review?

One year (2015-16)

4. What were the goals/objectives of the review?

The committee read literature, scheduled focus groups, and reached out to their tenure/promotion committee and their center for teaching and learning. They established a new focus on feedback for educators and learners to promote their development.

From Final Report: “In December 2015, the Provost formed our committee to ‘review Smith’s current teaching evaluation with a focus on how well it serves its diverse institutional functions and how consistent it is with current research and best practices regarding teaching evaluations’...During the spring semester of 2016, we read recent scholarship on student evaluations of teaching (SET) especially regarding questions of implicit bias and best practices; revisited previous course evaluation committee reports; analyzed other colleges’ instruments; and invited online and inperson feedback from chairs, directors, former members of T&P, the associate provost, students, and faculty members.”

5. What were the most useful resources (on or off campus) identified by the reviewers?

- See Stanford connection for student feedback sharing
- See MIT approach to post tenure assessment of teaching

- Stark, Philip B. and Richard Freishtat. "An Evaluation of Course Evaluations." ScienceOpen (September 24, 2014).
- Spooren, Pieter, Bert Brockx, and Dimitri Mortelmans. "On the Validity of Student Evaluation of Teaching: The State of the Art." Review Of Educational Research 83, no. 4 (December 1, 2013): 598642.

6. What data were collected by the reviewers?

None

7. What were the strengths of the former system for evaluating teaching? The weaknesses?

Faculty felt that the open-ended questions were useful

8. What were the major recommendations of the review?

- Importance of including students in this discussion and process
- Shift terminology from course evaluation to feedback
- Shift feedback process to the academic side of the hours

From FinalReport:

- "We recommend for adoption a new questionnaire (see attached) that asks for student reflection on learning outcomes and student experience of teaching methods, course structure, and format. In addition, it gives students an opportunity to say something about the course to other students for the purpose of making more informed course selection decisions.
- To address a number of issues that adversely affect students' attitudes about the course feedback process, we recommend changes in the way the questionnaire is implemented in terms of timing, eliminating repetition, and motivating students.
- Measuring teaching effectiveness calls for a broader range of tools and practices (e.g., observations, portfolios) to accompany student feedback about teaching, so we recommend that another committee be formed to tackle this challenge, especially for the institutional purpose of making decisions about employment and merit."

9. How were these recommendations communicated to administrators, faculty, and students?

10. What do you wish you had done differently with respect to the review?

11. What do you wish you knew at the beginning of the review-- that you know now?

- Think carefully about who your audience is for your report. What sells to what audience?
- Include at least two students in each committee (preferably upper-level students)
- Partner with Institutional Research as early as possible in the process— to effectively plan implementation. Note that Smith's Office of Institutional Research has 8-10 employees for 2100 students while W&M has 2 employees for 10,000 students.

Implementation-based questions

1. To what extent were the recommendations implemented?

The majority of the review committee recommendations were implemented

2. Who was involved in the implementation?

Floyd transitioned from chair of the ad hoc review committee to chair of the implementation committee. This committee included faculty and students and partnered closely with their Institutional Research office.

3. What was the timeline of the implementation?

Two years (2016-18). The original plan was to pilot the new feedback system in the first year (via Qualtrics) then pilot it to a broader audience in the second year (via Blue).

4. What were the goals/objectives of the implementation?

Goals

- To provide feedback to faculty to improve their teaching
- To provide data to the tenure/promotion committee (noting that student input is only one tool)
- To provide data to students on different classes (just data related to the final question on the form)

They implemented several changes to their feedback form:

- Changed name from “course evaluation” to “feedback”
- Reworded the first question on the form. OLD: Reflect on your attempts to learn in this course. Did you skip class? Did you do the reading? NEW: Reflect on the biggest takeaways from this course in terms of learning objectives and content.
- Changed the timing of feedback. OLD: three weeks before the end of the semester. NEW: immediately after the last day of classes.
- Changed the voice of survey to make it sound like faculty are asking students for feedback. OLD: Objective, impersonal wording made students question whether faculty even read their feedback. NEW: Wording adopted the voice of the faculty. e.g., “Tell me about the course structure.”

They also implemented other changes:

- Established a public database with results from the final question: What would you like to say about this course to a student who is considering taking it in the future? Note that faculty have the option to opt out of this.
- Student government association made feedback forms mandatory and assessed a \$25 fine on students who didn’t complete them.

5. How was the implementation assessed?

Not assessed

6. What were the strengths of the new system for evaluating teaching? The weaknesses?

Weakness: Too little turnaround time to process feedback in the spring semester

7. What do you wish you had done differently with respect to implementation?

8. What do you wish you knew at the beginning of the implementation-- that you know now?

9. How did they make decisions about allocating resources to develop and then to implement, a new system?

- They were told very explicitly by their Provost not to worry about funding and financial considerations

Extending work beyond feedback forms

1. What prompted the review?

The implementation committee felt that their review of how Smith evaluates teaching should extend beyond student feedback to include other aspects of evaluating teaching such as self-reflection and peer evaluation/

2. Who was involved in the review?

Floyd transitioned from chair of the implementation committee to chair of the committee on the Holistic Assessment of Teaching (HAT). HAT was composed of faculty only, with assistance from institutional research.

3. What was the timeline of the implementation?

One year (2018).

4. What were the goals/objectives of the implementation?

Goals

- To make recommendations to Provost on assessment.

They made several recommendations that were never implemented:

- Faculty assessments should include self-reflection and peer consultation (both observation and mentoring)
- Assessments should differ between pre- and post-tenure faculty—with concept that improving teaching should be a life-long goal, but often isn't
- Post tenure: a separate committee should determine if faculty are working on their teaching (assess 1/3 of post-tenure faculty per year); faculty should write reports on their teaching and professional development, what they are currently working on, including info on their mentoring and student feedback. These assessments should avoid Likert scales and be repeated every three years.

5. How was the implementation assessed?

6. What were the strengths of the new system for evaluating teaching? The weaknesses?

Weaknesses: HAT committee suggestions were too ambitious. After the changes made in the previous two years, there was no appetite for a pilot. Smith also had a new provost who was not motivated to make changes. Suggestions for giving teaching more weight in merit were viewed as too radical.

7. What do you wish you had done differently with respect to implementation?

8. What do you wish you knew at the beginning of the implementation-- that you know now?

9. How did they make decisions about allocating resources to develop and then to implement, a new system?

- They were told very explicitly by their Provost not to worry about funding and financial considerations

Additional info

Merit review at Smith: calculated out of 30 pts (10pts each for teaching, research, and service). Points and process not at all transparent. Faculty do not know their numbers. Pre-tenure faculty all receive a standard annual raise. Post-tenure faculty are scored level 0, 1, or 2 (2 = raise).

Case Study: University of Kansas

KU is a public R1 university in Lawrence, KS. It was founded in 1865.

Total 27,690 students have enrolled into University of Kansas where 19,596 students have enroll into undergraduate programs and 8,094 students for graduate programs.

By gender, 12,997 male and 14,693 female students are attending KU; by race/ethnicity, 19,151 White, 1,150 Black and 1,287 Asian students out of total 27,690 are attending the University of Kansas

Student Demographics (Fall 2019)

Group	Percent of Undergraduates
White	70.7%
Hispanic/Latino	7.9%
Non-Resident Alien	6.2%
Asian	4.7%
Black or African American	4.2%
American Indian or Alaska Native	0.4%
Native Hawaiian or Other Pacific Islander	0.1%
Ethnicity Unknown	5.8%

This is a report from Andrea “Dea” Follmer Greenhoot, Director and Gautt Teaching Scholar, Center for Teaching Excellence And Professor, Department of Psychology at Kansas:

Our effort at KU has been undertaken from many directions, and we have a lot going on, but we are not quite at the point where the university has “decided” to formally change its evaluation procedures. Indeed, it’s not even clear that a formal policy change is needed at KU, because we already require that teaching evaluations draw on three sources (students, peers, and instructor) and focus on multiple dimensions of teaching practice. The issue has been that this policy is not implemented much at all in practice- most evaluations still focus almost exclusively on student ratings. So KU’s work on this started with the Benchmarks framework we developed out of the [Center for Teaching Excellence]. The framework takes the form of a rubric that identifies

multiple dimensions of teaching and draws on evidence from multiple sources. We developed it several years ago as a tool for departments to integrate information collected from multiple sources (peers, instructor and students)-basically to close that gap between policy and practice. Here is an outline of steps taken at KU since the initial development:

- 1- **2016-2017.** Refined framework and rubric by gathering systematic input from various stakeholders across the university (including ambassadors from each academic department and department chairs), and some department level piloting.
- 2- **2017-2022 .** In 2017 we received an NSF grant with collaborating institutions to launch a project (TEval- see www.TEval.net) that supports STEM departments on our campuses in adapting and implementing the framework for some form of formative and/or summative evaluation (all with the goal of ultimately reshaping evaluation of teaching for P&T. At this time, the Deans of the college of liberal arts and sciences and of engineering, along with the vice provost for faculty development, all committed to fostering discussion of this topic and to connecting us to school and university level evaluation committees. Each year, each campus has brought more departments into the project, with each serving as an incubator for designing and testing processes for improved teaching evaluation.
- 3- **2019.** Faculty senate charged a committee, co-led by my colleague (and TEval co-PI at KU) Doug Ward, to investigate best practices for online student rating systems. Their recommendations included one to **de-emphasize** the focus on student ratings. [see attached report]
- 4- **2019-2020:** The Vice Provost for faculty development launched a task force to reconsider the role and methods of gathering student ratings all together, for the purposes of P&T. The goal is for the task force to have recommendations by December 2020. Meanwhile, CTE and the VP of FD collaborated with HR to have our rubric programmed into the online HR annual evaluation system (MyTalent) as a tool for the evaluation of lecturers and teaching professors (can tell you more about that if you want).

So at the moment, there is no top down mandate to adopt our framework-indeed, we feel that would derail our efforts at this point bc departments have a high level of investment in what they are doing. But we are at a point where we have been able to catalyze and support converging efforts from all levels of the system, and I think we are close to a university-wide change in expectations for evaluation practice that would align with the Benchmarks framework.

Moreover, this is all happening in a context where there's a growing national conversation about this topic. For instance, TEval just collaborated with the AAU and the National Academies to host a meeting of leaders from universities across the country to begin to frame the national dialogue on this topic, and the proceedings are now available here.

<https://www.nap.edu/catalog/25685/recognizing-and-evaluating-science-teaching-in-higher-education-proceedings-of>

University of Oregon

Location: Eugene, OR

Student Population

Undergraduates- 19,101

Graduates- 3,543

Private v. Public: Public

Year Founded: 1876

Student Demographics:

53.7% Women

8,842% Men

0.6% American Indian/Alaskan Native

6.7% Asian

2.5% Black/African American

14.2% Hispanic/Latino

9.1% Multi-race (not Hispanic/Latino)

0.5% Native Hawaiian/Pacific Islander

64.6% White

1.7% Unknown

International Students 9.8% from 83 Countries

First Year Students Returning 85%

Students Graduating within 4 Years 52.9%

Point people: Sierra Dawson and Bill Harbaugh

Phone Interview with Sierra Dawson, Associate Vice Provost for Academic Affairs:

Review-based questions

What prompted the review?

The conversation was started because a professor brought a specific issue before the Faculty Senate. They accused a student of academic misconduct, however, because of how the judicial process occurred that student was able to fill out a student evaluation for the course/professor. The professor was worried about how a student would retaliate and how that would impact the evaluation of their teaching. This sparked a larger conversation about how Oregon conducted teaching evaluations.

Who was involved in the review?

Office of the Provost and the Faculty Senate

A task force was created in 2017 to make initial recommendations. In 2018 the Faculty Senate created a standing committee to address these issues.

What was the timeline of the review?

Spring 2017:

-Senate Motion to create task force

Fall 2017:

-Meetings with stakeholders: student groups (graduate and undergraduate), staff, faculty etc.

Winter 2018:

- Site visits from Emily Miller (Associate Vice President for policy at the Association of American Universities) and Noah Finkelstein (Director of the Center for STEM Learning at the University of Colorado Boulder)
- Initial pilot of midterm student experience survey
- Initial pilot of student experience survey
- Initial pilot of faculty instructor reflection

Spring 2018:

- Update surveys and reflections
- Townhalls with students (undergraduate and graduate), unit heads, and faculty

Summer 2018:

- Faculty teaching over the summer pilot the Midterm Student Experience Survey, End of Term Student Experience Survey, and the 10-Minute Instructor Reflection

Fall 2018:

- Create cohort of departments into pilot subgroups
- Provide guidance for updating

Winter 2019:

- Continues to pilot new tools

Spring 2019:

- Continues to pilot new tools

Fall 2019:

- First semester of changes made

What were the goals/objectives of the review?

- Fair and transparent process
- Evaluations would be conducted against criteria of excellence (What was University of Oregon's criteria for teaching excellence)
- Gathered input from students, faculty, and staff

What were the most useful resources (on or off campus) identified by the reviewers?

- Office of the Provost's Teaching Academy

What data were collected by the reviewers?

- Read current literature
- Reached out to peer institutions (in particular USC)
- Reached out to stakeholders to get feedback on process
- Got feedback after pilot studies

What were the strengths of the former system for evaluating teaching? The weaknesses?

Strengths:

- incorporated peer and student review
- Fast and efficient

Weaknesses:

- Gave precedence to student voice
- Used the means of student evaluations to rate professors
- Did not have clear guidelines for what Oregon considered teaching excellence

What were the major recommendations of the review?

- Set clear definition of what teaching excellence is
- Understand what is actually being evaluated
- Place greater emphasis on peer and individual evaluation

How were these recommendations communicated to administrators, faculty, and students?

- Emails
- Met with Students
- Met with every department/division

What do you wish you had done differently with respect to the review?

- N/A

What do you wish you knew at the beginning of the review-- that you know now?

- Creating an evaluation system
- What is the University of Oregon evaluating?
- Long and laborious process

Implementation-based questions

To what extent were the recommendations implemented?

- The full set of recommendations were rolled out in Spring Fall 2019
- Beginning to start feedback campaign

Who was involved in the implementation?

- Faculty senate
- Task force
- Departments
- Provost's office

What was the timeline of the implementation?

- Process started in Spring 2017 and just had full scale roll out in Fall 2019

What were the goals/objectives of the implementation?

- Communication
- Transparency

How was the implementation assessed?

-Conducting assessment currently

What were the strengths of the new system for evaluating teaching? The weaknesses?

-Emphasizes personal and peer reflection

-Allows professors to reflect on teaching before the tenure process

What do you wish you had done differently with respect to implementation?

-Communication is key!

What do you wish you knew at the beginning of the implementation-- that you know now?

-N/A

Case Study: USC

USC is a private R1 university in downtown Los Angeles, CA. It was founded in 1879.

There are currently 20,500 undergraduates and 28,500 graduate and professional students enrolled at USC. There are 22 professional schools at USC beyond “Letters, Arts, and Sciences.” It is clearly quite a different institution than WM; it’s student demographics reflect this, as well.

Student Demographics (Fall 2019)

Asians	16.9%
Black/African-American	5.3%
Hispanic	14.6%
White/Caucasian	29%
International	25.4%
Other	8.8%

USC began the process of reviewing it’s teaching evaluation process in 2014. This review was precipitated by a survey carried out by their faculty senate that demonstrated a great deal of dissatisfaction among the faculty surrounding tenure and promotion. Those concerns encompassed most aspects of the tenuring process, but what caught the attention of the Senate and Provost at the time were concerns about the biases and inconsistencies of teaching evaluation. The Provost, Michael Quick, was strongly supportive of transforming teaching and its evaluation at USC, and so supported the development of a Center for Excellence in Teaching, directed by Assoc Vice Provost Ginger Clark. <http://cet.usc.edu/about/overview/>

The first, and vital, step undertaken by this Center was coming up with a definition of excellence <http://cet.usc.edu/about/usc-definition-of-excellence-in-teaching/>. “Faculty experts in

education and curriculum from across disciplines were consulted to help develop a plan,” according to their Excellence in Teaching Initiative, and further charged with implementing the plan as well.

Implementing this plan had two primary dimensions. First, it required a serious commitment of resources to the Center. Second, it required buy in and full participation by all of the schools and programs across the (vast!) USC campus. The resource commitment has been substantial; the CET began with two full time Instructional Designers, and later a third was added. These designers are responsible for Arts and Humanities, Social Sciences, and STEM respectively. Getting all of these resources in place took roughly a year and a half.

Implementation included planning at the outset. That is to say, each division and school were expected to modify the definition of excellence proposed by the CET to suit its own mission, and to hold itself accountable for fulfilling their own definition of excellence accordingly.

Implementing the plan is STILL ongoing; not all teaching is yet subject to peer review, but that is the plan. Peer review is (according to the PowerPoint New USC Model of

Teaching Excellence) the primary means by which “Teaching Effectiveness” will be evaluated going forward.

Assoc Provost Clark pointed to a number of critical points, as well as obstacles in implementing these changes. First, it is imperative that good teaching be rewarded through grants and even incentives like course releases and leaves. Further, it needs to be clear that the research is really not there to demonstrate that peer review is better than student evaluation – it simply hasn’t been carried out at the scale that would permit that kind of finding to be made. Finally, she noted that buy in needs to come from the bottom up, and not simply through governance structures. Frequent workshops, pilot programs, and targeted faculty opportunities to work with the Center would improve what (for USC) was a cumbersome process managed primarily by governance.

Vanderbilt University

Location: Nashville, TN

Student Population

Undergraduates- 6,871

Graduates- 5,716

Private v. Public: Private

Year Founded: 1873

Student Demographics:

50.5% Women

49.5% Men

0.5% American Indian/Alaskan Native

13.4% Asian

9.9% Black/African American

9.9% Hispanic/Latino

5.3% Multi-race (not Hispanic/Latino)

0.2% Native Hawaiian/Pacific Islander

55.4% White

5.5% Unknown

International Students 7.1% from 47 Countries

First Year Students Returning 97%

Students Graduating within 4 Years 85.9%

Point people: Leona Schauble and Mark Schoenfield

Phone Interview with Leona Schauble (Co-Chair of Committee):

Review-based questions

What prompted the review?

When Vanderbilt moved away from pencil and paper evaluation the percentage of students participating in the process dropped by 25%. The low response rate created a space where the course evaluation were not useful anymore.

Who was involved in the review?

The changes were based on the work of four committees with faculty and student representation, and were vetted by the faculties of the four undergraduate schools. Initial recommendations were made by the Committee to Reevaluate Evaluations in October 2014.

Feedback was gathered from faculty within each of the four undergraduate schools in Spring 2014 and incorporated into the design of the evaluation questions by the Teaching Evaluation Finalization Task Force in Summer 2015.

What was the timeline of the review?

Fall 2014: Initial recommendations were made by the Committee to Reevaluate Evaluations

Spring 2014: Feedback was gathered from the undergraduate schools

Summer 2015: Incorporated feedback into design

Spring 2016: Recommendations for implementation

Fall 2016: Roll out new system

What were the goals/objectives of the review?

- To provide student feedback to the faculty on the pedagogy and delivery of instruction.
- To inform departments and administration about faculty teaching in tenure, promotion, and reappointment.
- Student evaluations should provide students with data to inform course and instructor selection.

What were the most useful resources (on or off campus) identified by the reviewers?

- Center for Teaching

What data were collected by the reviewers?

- At what point do you want to ask for feedback?
- At what points will that feedback important?
- Collected information from other institutions like Vanderbilt
- Literature review of the literature with focus on bias, how students respond to sensitive topics and class size

What were the strengths of the former system for evaluating teaching? The weaknesses?

Strengths:

- Fast

Weaknesses:

- No time in class for completion
- Wording of questions
- Results were compared to department means for similar courses, so it had a major impact on tenure

What were the major recommendations of the review?

- The evaluation questionnaire was revised to provide clearer questions for students and more useful feedback for faculty and administrators
- The revised evaluation questionnaire is accessed via mobile devices
- Faculty are encouraged to address the evaluation process with students, through classroom discussion and/or the syllabus
- Faculty are encouraged to allot ~20 minutes of class time for students to complete evaluations
- There will be ongoing assessment

How were these recommendations communicated to administrators, faculty, and students?

- Focus groups
- Spent time talking to individual departments and schools

What do you wish you had done differently with respect to the review?

- Learn how to convey the importance of changing student evaluations to faculty members

What do you wish you knew at the beginning of the review-- that you know now?

- More talking and more feed-back

Implementation-based questions

To what extent were the recommendations implemented?

-Not all recommendations were implemented

Who was involved in the implementation?

-Had a separate implementation committee who took the recommendations from the task force.

What was the timeline of the implementation?

-One semester

-Overall process was two years

What were the goals/objectives of the implementation?

-Transparency

How was the implementation assessed?

-Feedback from faculty and students

What were the strengths of the new system for evaluating teaching? The weaknesses?

-Clear and concise questions

-Easily accessible

-Stresses importance to students

What do you wish you had done differently with respect to implementation?

-Learn how to convey the importance of changing student evaluations to faculty members

What do you wish you knew at the beginning of the implementation-- that you know now?

-N/A