Supplemental Material

CBE—Life Sciences Education

Krishnan *et al*.

SUPPLEMENTAL MATERIALS FOR

Guides to Advance Teaching Evaluation (GATEs): A Resource for STEM Departments Planning Robust and Equitable Evaluation Practices

Sandhya Krishnan, Jessica Gehrtz, Paula P. Lemons, Erin L. Dolan, Peggy Brickman, Tessa C. Andrews

Guides to Advance Teaching Evaluation (GATEs) in STEM Departments

This document provides actionable guidance for the long-term development of departmental practices for robust and equitable teaching evaluation.

DeLTA
UNIVERSITY OF GEORGIA
Scientists Engaged in Education Research

Three voices inform teaching evaluation:

- **Peer voice** involves gathering data from peers about teaching and learning occurring in an instructor's class. This document focuses on peer observation.
- **Student voice** involves gathering data from students about their learning and perceptions. This document focuses on mandatory student evaluations AND other sources of data from students.
- **Self voice** involves a written narrative documenting a systematic self-reflection process.

For each voice, robust and equitable evaluation is:

- **Structured**: Evaluation that is structured ensures fairness and minimizes bias. Structure involves processes that are formalized (i.e., written down) and fair, training and support for faculty, and collective decision-making among department members to develop and enact policies and practices.
- **Reliable**: Evaluation that is reliable is informed by multiple sources of meaningful and trustworthy evidence.
- **Longitudinal**: Evaluation that is longitudinal is able to document improvement overtime and provide feedback to faculty about strengths and room for improvement.

The Guide for each voice has three components. These Guides:

- Specify **Target Practices,** which are long-term goals departments can work toward. These were developed based on research and successful practices at research-intensive institutions, and are formatted as a self-assessment.
- Characterize common **Starting Places** departments may be when they begin considering teaching evaluation practices.
- Provide ideas for **Starting Strong and Engaging Efficiently**, including quick-start ideas, "bundles" of target practices that may be efficiently accomplished together, and links to outside resources.

Contact Tessa Andrews (tandrews@uga.edu) for more information. These guides were created by the DeLTA Project at the University of Georgia with support from the National Science Foundation (DUE 1821023). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Peer Voice Target Practices

Peer voice involves gathering data from peers about teaching and learning observable in class. Peer observation incorporates multiple steps:

- Pre-observation meeting to discuss lessons to be observed
- Collection & review of class materials (e.g., syllabi, exams, homework, slides, handouts)
- Observation of lessons
- Post-observation meeting to hear instructor reflections, debrief, and provide feedback

		Peer Voice Target Practices: What is your status and what actions will you take?	Not right now	Want to work on it	Working on it	Fully in place
Structured	1	Department uses a formal observation form to guide what is observed and which other data are collected (e.g., class materials, assessments, pre-observation meeting). Forms may be adopted or adapted from other departments.	☐	☐	☐	☐
	2	Department has a formal template for writing a report based on peer review, potentially distinguishing between formative and summative review.	☐	☐	☐	☐
	3	Department uses formal processes or criteria to select peer observer(s) for all instructors.	☐	☐	☐	☐
	4	Department enacts policy about the number of peer observations & observers during a review period and/or across review periods.	☐	☐	☐	☐
	5	Department designates a coordinator, leader, or committee to carry out and refine peer observation practices.	☐	☐	☐	☐
	6	Department has a process for allocating and recognizing workload related to coordinating and conducting observations.	☐	☐	☐	☐
	7	Department periodically discusses and improves peer evaluation practices to maximize utility to instructors and the department.	☐	☐	☐	☐
	8	Department provides or arranges formal training about the departmental peer review process for peer observers.	☐	☐	☐	☐
Reliable	9	Department relies on multiple observations for all instructors, such as using multiple observers, observing multiple lessons, and/or observing multiple courses.	☐	☐	☐	☐
	10	Department specifies which class materials (e.g., syllabi, exams, homework, slides, handouts) are collected and evaluated as part of peer observation.	☐	☐	☐	☐
	11	Department expects observers to talk with instructors to properly contextualize observations and review of materials. This might include discussing course goals, lesson goals, class structure, and students.	☐	☐	☐	☐
Longitudinal	12	Department conducts peer observation over multiple time points in a review period for all instructors to document teaching improvements.	☐	☐	☐	☐
	13	Department ensures that the peer observation process provides feedback to instructors via follow-up discussion that covers strengths and areas for improvement.	☐	☐	☐	☐

ABSENT	Department does <u>not</u> use peer evaluation to inform teaching evaluation.

BITS & PIECES	Peer evaluation occurs without any explicit departmental policies or practices.
	Department relies on just one source of evidence for peer observation, such as a single observation of a single lesson.
	Department does <u>not</u> expect peer observation to be conducted more than once.

CLOSER TO COHESION	Department enacts peer observation process that falls back on historical precedent or is idiosyncratic to each observer and candidate regarding: • The logistics of peer observation (e.g., selection of observers, number of observers, when observed) • The observation criteria • The report produced by observer(s) Department provides some coordination, possibly inconsistent, of peer observations.
	Department expects more than one source of evidence for peer observation. For example: • More than one observer • More than one lesson observed in the same course • More than one course • Collection and evaluation of class materials • Conversations between candidates and observers
	Department documents teaching improvements for *some* candidates by conducting peer observation over multiple time points. For example, this may only occur for: • Faculty with majority teaching EFT • Junior faculty • Faculty with consistently low student evaluations • Faculty with peer observations that indicate areas of concern • Other: _____ Department does <u>not</u> ensure that the peer observation process provides feedback on strengths and suggestions for growth to faculty.

TARGET PRACTICES

Where is your department starting?

A, B, and C are common starting places for departments working to reform how they use **PEER VOICE** in teaching evaluation. Reflecting on current practices can illuminate what target practices are a good next step. **Does A, B, or C best align with the current practices in your department?**

Starting Strong and Engaging Efficiently with the Peer Voice

Based on experiences with STEM departments, we suggest potential entry points for expanding target practices. We also provide "bundles" to highlight how work on one target practice can be leveraged to achieve other target practices.

Legend
Colors refer to Target Practices that are:
Structured
Reliable
Longitudinal

Two Quick Start Ideas

- Convene a committee (#5)
- Charge committee with developing or adapting observation form** (#1)
- Consider workload equity (#6)

- Examine observation forms** developed by other departments (#1)
- Pilot adapted observation forms with willing faculty to start discussion about peer evaluation practices (#7)
- Determine how to provide feedback about teaching strengths and areas for improvement (#13)

**Go to: https://tinyurl.com/GATEsExtraResources for links to example peer observation forms.
See sheet labeled "Peer voice resources"

Two Potential Bundles

When developing or adapting a peer observation form** (#1), determine:

- …what class materials will be collected (#10)
- …how feedback will be provided to candidates (#13)
- …how observers talk to instructors to get a sense of the big picture (#11)
- …how peer evaluation results will be relayed to the department (#2)

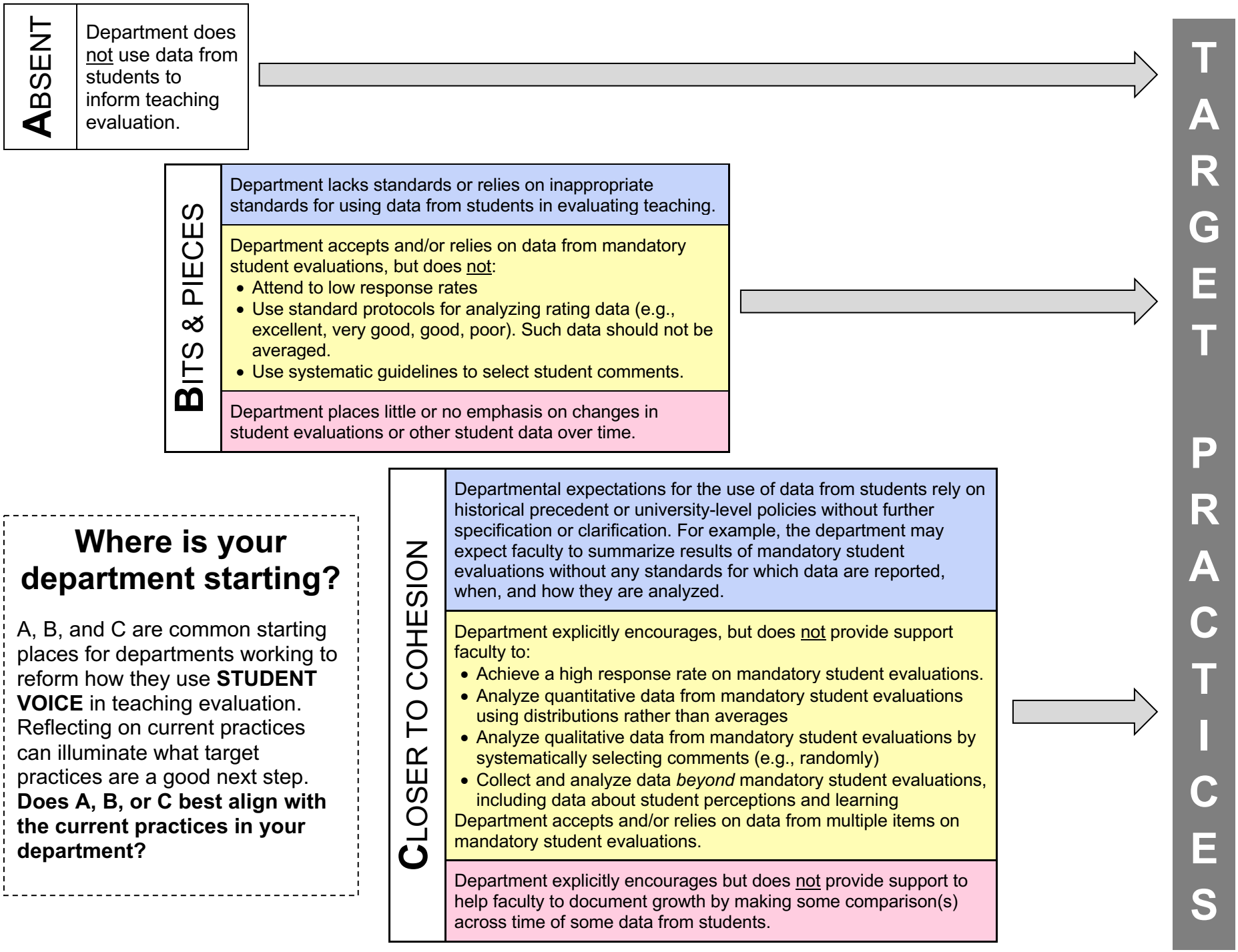
Decisions about how to implement peer observation can be made at the same time, including:

- …how observers are selected (#3)
- …how many observers are used (#4)
- …how many observations occur (#9)
- …when observations occur (#12)

Student Voice Target Practices

Student voice involves gathering data from students about their learning and perceptions. Mandatory end-of-course evaluations are a common source of student voice. Evidence from students should go beyond course evaluations. This could include, but is not limited to: data on learning, grade anomalies or opportunity gaps, mid-term evaluations or classroom interviews, research-based assessment results, instructor-created surveys.

		Student Voice Target Practices: What is your status and what actions will you take?	Not right now	Want to work on it	Working on it	Fully in place
Structured	1	Department has formal standards for how and when instructors collect, analyze, and report student data (e.g., response rate expectation, standard quantitative and qualitative analysis).	☐	☐	☐	☐
	2	Department makes appropriate distinctions in their expectations about student data for different review periods (e.g., annual review, 3rd year review, promotions) and different levels of teaching experience with a given course.	☐	☐	☐	☐
	3	Department periodically discusses and improves expectations for collecting and analyzing data from students to maximize utility to instructors and the department.	☐	☐	☐	☐
	4	Department provides or arranges formal training, or other support, for instructors about collecting and analyzing student data, including achieving high response rates, analyzing quantitative and qualitative data systematically and appropriately, gathering data beyond mandatory evaluations, and making comparisons across time.	☐	☐	☐	☐
Reliable	5	Department expects instructors to do everything they can to achieve high response rates on mandatory student evaluations (e.g., course credit offered, class time set aside).	☐	☐	☐	☐
	6	Department recognizes known biases, such as bias against women, minoritized groups, and large class size, and limits comparisons of mandatory student evaluations between instructors.	☐	☐	☐	☐
	7	Department specifies that quantitative questions on mandatory student evaluations be analyzed as distributions of scores, rather than averages. Because quantitative questions often use an ordinal rating scale (excellent, very good, good, poor), average scores and standard deviations are inappropriate. We cannot assume the points on ordinal scales are equidistant.	☐	☐	☐	☐
	8	Department specifies which set of quantitative student evaluation questions are used for each review period (e.g., annual, promotion).	☐	☐	☐	☐
	9	Department specifies that student comments on mandatory evaluations be systematically examined to determine teaching strengths and room for improvement.	☐	☐	☐	☐
	10	Department expects instructors to collect, analyze, and interpret some data *beyond* mandatory student evaluations.	☐	☐	☐	☐
Longitudinal	11	Department expects instructors to document change (or consistently exemplary results) by comparing data from students across multiple timepoints.	☐	☐	☐	☐



Where is your department starting?

A, B, and C are common starting places for departments working to reform how they use **STUDENT VOICE** in teaching evaluation. Reflecting on current practices can illuminate what target practices are a good next step. **Does A, B, or C best align with the current practices in your department?**

Starting Strong and Engaging Efficiently with Student Voice

Based on experiences with STEM departments, we suggest potential entry points for expanding target practices. We also provide "bundles" to highlight how work on one target practice can be leveraged to achieve other target practices.

Legend
Colors refer to Target Practices that are:
Structured
Reliable
Longitudinal

Two Quick Start Ideas

Explore solutions to increase response rates** on mandatory student evaluations (#5)

Learn about and determine how to account for known biases** (#6)

**Go to: https://tinyurl.com/GATEsExtraResources for info about:
(1) increasing evaluation response rate
(2) bias on student evals
(3) analyzing quantitative data
See sheet labeled "Student voice resources"

Two Potential Bundles

Set standards (#1) for mandatory student evaluations about how:

- …instructors attempt various options to increase response rate (#5)
- ….a standard set of quantitative items (#8) are analyzed** with distributions rather than means (#7)
- …student comments for open response items are systematically analyzed (#9)

Set expectations for faculty to gather data *beyond* mandatory student evaluations (#10).

- Support faculty in gathering and analyzing these data through training (#4)
- Support faculty to compare these data over time to document growth (#11)

Guide to Self Voice Practices

Self voice involves a written narrative documenting the self-reflection process. Self-reflection helps faculty continuously improve their teaching by critically considering evidence. Formal documentation of this process can provide valuable information for evaluating teaching, and in particular can document aspects of teaching that are not obvious to students or observers.

- The process of self-reflection involves the **collection of evidence and/or systematic observation**, and **analysis** of the evidence/observations to answer a question.
- The written narrative documents this process such that faculty **reflect** on the findings to make sense of them and plan next steps.

		Self Voice Target Practices: What is your status and what actions will you take?	Not right now	Want to work on it	Working on it	Fully in place
Structured	1	Department uses a formal self-reflection form to guide the scope and content of written self-reflection narratives, including standards for what constitutes evidence-based self-reflection. Forms may be adopted or adapted from other departments.	☐	☐	☐	☐
	2	Department periodically discusses and improves standards for written teaching reflections to maximize utility to instructors and the department.	☐	☐	☐	☐
	3	Department provides or arranges formal training, or other support, for instructors about the self-reflection process and to help instructors meet departmental expectations for documenting self-reflection.	☐	☐	☐	☐
Reliable	4	Department expects instructors to engage in a self-reflection process, and written documentation thereof, that is focused on tackling teaching challenges (e.g., concerns raised in student evaluations or peer observation, student learning difficulties, lack of engagement).	☐	☐	☐	☐
	5	Department expects the self-reflection process, and written documentation thereof, to rely on the systematic analysis of evidence about student learning and experiences.	☐	☐	☐	☐
	6	Department expectations for self-reflection consider the experience level of instructors. For example, instructors new to a course or teaching may primarily rely on informal sources of data (e.g., notes, brief written feedback from students), whereas more experienced instructors rely on formal sources of data (e.g., assessment data) and systematic observation (e.g., feedback from trained peers).	☐	☐	☐	☐
Longitudinal	7	Department expects that written reflections discuss how instructors have built on prior self-reflections, including the outcomes of planned improvements and innovations.	☐	☐	☐	☐
	8	Department expects that written reflections discuss efforts to grow and learn as educators. This can include learning from both successes and failures.	☐	☐	☐	☐

ABSENT	Department does not use written self-reflections to inform teaching evaluation.

BITS & PIECES	Department lacks standards for written teaching reflection.
	Department asks faculty to submit written reflection on teaching activities but does not expect faculty to reflect on evidence or systematic observations.
	Department does not expect written descriptions to address change over time.

Where is your department starting?

A, B, and C are common starting places for departments working to reform how they use **SELF VOICE** in teaching evaluation. Reflecting on current practices can illuminate what target practices are a good next step. **Does A, B, or C best align with the current practices in your department?**

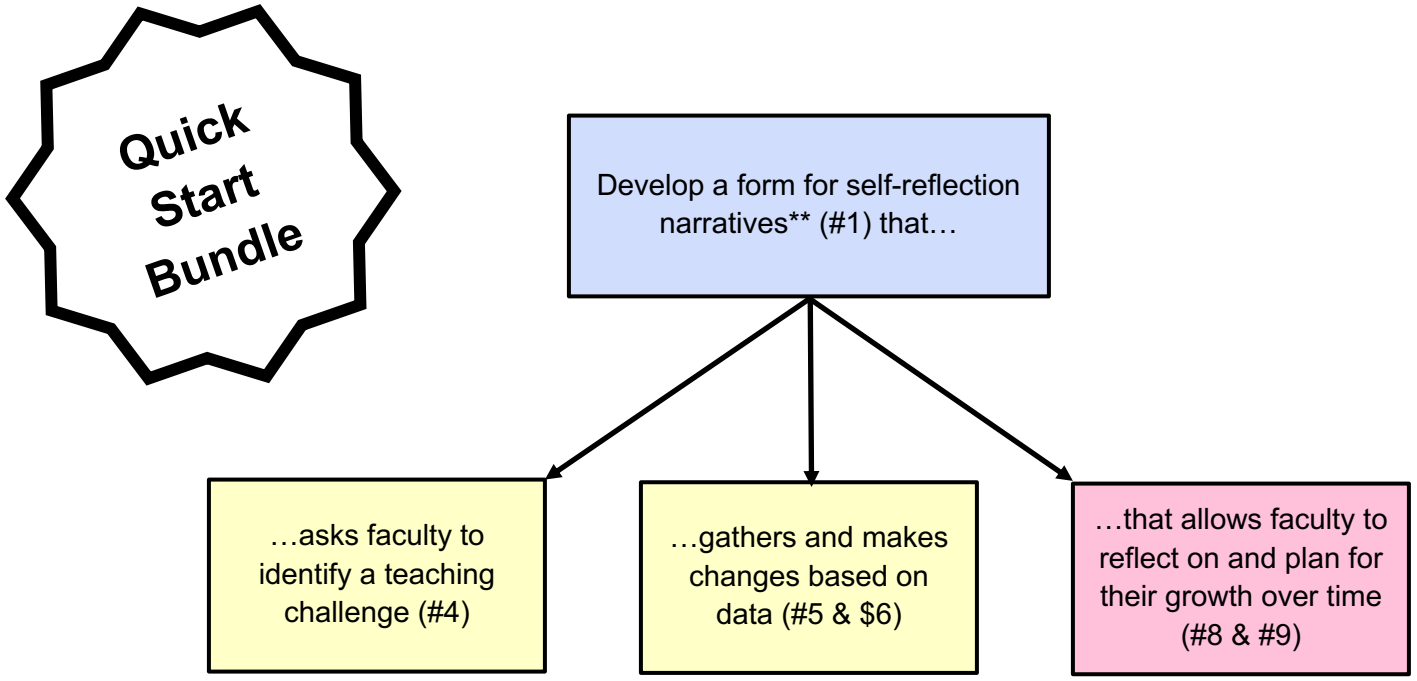
CLOSER TO COHESION	Department suggests, but does not mandate, standards for written teaching reflection.
	Department explicitly encourages but does not expect nor support faculty to: • Write reflections that consider some outside evidence or observations, such as concerns raised in mandatory course evaluations or peer observation. • Write reflections that describe how the instructor used evidence or observations to inform decisions about what and how to change.
	Department explicitly encourages but does not expect nor support faculty to write reflections that describe changes over multiple semesters of teaching, including innovations and improvements.

TARGET PRACTICES

Starting Strong and Engaging Efficiently with Self Voice

Based on experiences with STEM departments, we suggest a potential entry point "bundle" to highlight how work on one target practice can be leveraged to achieve other target practices.

Legend
Colors refer to Target Practices that are:
Structured
Reliable
Longitudinal



**Go to: https://tinyurl.com/GATEsExtraResources for examples of self-reflection forms and rubrics to help evaluate self-reflections. See third sheet labeled "Self-voice resources"

APPENDIX B: SUPPORTING EVIDENCE FOR TARGET PRACTICES

If the links are not accessible in this version, please access a live version here: https://tinyurl.com/UGADeLTAGATEsAppendixB

This document provides the supporting evidence for each target practices across peer, student, and self voice. This document was not designed to serve as a resource to departments or department leaders. It was designed to organize and share the scholarly work and existing reform efforts that lend support to each target practice. The links will become out of date. Additionally, as more scholarship is undertaken and more reform efforts take place, we will learn more about which teaching evaluation practices are important. We encourage readers to view the evidence presented here as the best available evidence in 2021.

PEER VOICE

Structured: Ensures fairness & minimizes bias through formalized processes, collective decision-making, and training

1. Department uses a formal observation form to guide what is observed and which other data are collected (e.g., class materials, assessments, pre-observation meeting). Forms may be adopted or adapted from other departments.
    a. Rationale: A formal observation form helps to standardize what observers pay attention to, externalizes what is valued as effective teaching, and can result in evidence that facilitates comparisons across time.
    b. Support includes recommendations from multiple universities:
        i. A common form to guide peer observation is common across multiple universities pursuing peer voice as part of teaching evaluation. See examples of peer observation forms from nine different institutions in the outside resources document linked to the guides: Resources

2. Department has a formal template for writing a report based on peer review, potentially distinguishing between formative and summative review.
    a. Rationale: A formal template for writing a report results in evidence from peer evaluation being synthesized and summarized in a similar way for each observation. This standardizes the evaluation that faculty experience and facilitates comparisons across time.
    b. Support includes recommendations from multiple universities:
        i. University of Oregon: In the 2019 Peer Review Guide form titled, "How Well is Peer Review Working in Your Unit?"on pg 3 titled "Solution Two: Template for peer review": "By creating a template for the output of a Peer Review, the unit/department, school/college and university personnel committees can expect consistent, robust reports that provide information that is valuable for both a) continual course improvement and b) evaluation of teaching excellence"
        ii. University of Colorado – Boulder: The webpage with Tools for Teaching Evaluation, under the section "Letter Writing Guides for Peer Observation"

includes an example from one department. This department opted to have a standard letter template rather than a standard peer observation protocol.
        iii. University of Arizona: Recommends guiding questions for writing a report following peer review

3. Department uses formal processes or criteria to select peer observer(s) for all instructors.
    a. Rationale: Using formal criteria ensures consistency across faculty, which is more equitable. Developing these processes/criteria requires conversations about what attributes are important for observers, which helps externalize values about teaching expertise.
    b. Support includes recommendations from a university and peer-reviewed research:
        i. University of Colorado - Boulder: Asks departments to describe a process for who will appoint faculty to conduct observations in the peer evaluation plan template
        ii. Gormally et al., 2017: Effective feedback requires the recipient to consider the reviewer as a reputable source. By making clear the expectations for peer observers, faculty will be more likely to trust that observers have appropriate expertise to provide effective feedback.

4. Department enacts policy about the number of peer observations & observers during a review period and/or across review periods.
    a. Rationale: Enacting a formal policy ensures consistency across faculty, which is more equitable. Establishing this policy also requires discussion and some degree of consensus about what review is necessary and useful for faculty at different career stages and in different positions.
    b. Support includes recommendations from one university:
        i. University of Colorado - Boulder articulates clearly that faculty of different ranks have different criteria for the number of evaluations per review period and a rationale for those decisions on pg. 4 of their peer evaluation plan template.

5. Department designates a coordinator, leader, or committee to carry out and refine peer observation practices..
    a. Rationale: Planning and coordinating peer observations requires time each semester and communication between leadership, review committees, observers, and faculty to be observed. It also requires coordinating training for peer observers. The time investment required is higher when departments are first adopting peer observation or modifying their approaches to peer observation. ASupport includes recommendations from multiple universities and our own experiences in the UGA DeLTA project:
        i. University of Colorado – Boulder: Template Peer Evaluation Plan for a Department. This will download a document. This template asks departments to describe the process for identifying faculty to conduct observations. The template specifically suggests a committee or informal group oversee the process by meeting regularly (at least twice per year).
        ii. University of Oregon: Recommends a coordinator to schedule peer observations, who then makes clear the schedule for observations in their Peer

Evaluation Guide for Departments: "Each unit/department might appoint a faculty or staff member as Peer Review Coordinator to oversee the scheduling of all peer reviews for the year. The peer review schedule could be shared with all faculty by week 1 of Fall term so that adjustments can be made if needed/requested"

iii. University of Georgia: We observed that departments who created a leadership position or committee responsible for peer observation or teaching evaluation more broadly made are making considerably more progress in working toward target practices.

6. Department has a process for allocating and recognizing workload related to coordinating and conducting observations.
    a. Rationale: Planning and coordinating peer observations requires time each semester and communication between leadership, review committees, observers, and faculty to be observed. Being intentional about how this workload is allocated and recognized as departmental service will help peer observation practices be sustainable over time.
    b. Support includes recommendations from multiple universities and peer-reviewed research:
        i. University of Colorado - Boulder: Peer Evaluation Plans template first asks departments to describe the process for which departmental bodies will appoint faculty to conduct observations. Then it specifically mentions that this committee/body is responsible for balancing workload: "The committee [or insert applicable body(ies)] will also appoint faculty to conduct observations. In most cases, the observer will be familiar with the course and/or course content, come from the same program, and be senior in rank to the observed faculty member. Faculty members may notify the committee if they prefer not to be observed by a specific colleague. However, in order to balance workload, schedules, etc., the [insert appropriate body(ies) as inserted above] will have final say in the selection of observers."
        ii. University of Oregon: Recognized that workload needs to be considered for peer observation, when they wrote in the Peer Review Guide for Departments that: "Each unit could identify and train a group of faculty to serve as peer reviewers. Participation would count as important unit/department level service, and typically requires 4-6 hours of service per faculty review. The unit/department could either train all faculty, or only the subset of faculty who will perform all peer reviews for the year. Faculty who will serve as reviewers should be identified at the start of each academic year"
        iii. There is growing recognition in research literature about inequities in faculty work by gender and race. These inequities take many forms. Faculty with particular social identities may feel more responsibility to engage in particular work. Faculty with certain identities may also have to invest additional time and effort into the same work because of social expectations and biases based on their identity. Finally, departments paying attention to equity in workload can create more equitable distribution of work. Collectively, these and many other papers suggest that it is important for departments to recognize workload related to teaching evaluation:

1. Baez, 2000: Qualitative study of how faculty of colour engage in service. Faculty of colour felt an increased responsibility to engage in service, particularly to aid students from traditionally less represented backgrounds that they did not believe was felt by white colleagues. In addition, they felt that this work was not recognized for promotion in the department.
2. Griffin & Reddick, 2011: Male and female Black faculty experienced race- and gender-related expectations regarding their mentorship of students. These pressures increase the burden of mentorship compared to white faculty.
3. O'Meara et al., 2017: There are gendered differences in faculty workload, with women doing more mentoring and teaching and receiving more activity requests
4. O'Meara et al., 2018: As departments learn how to attend to equity issues in workload in the department, faculty were more likely to report practices within the department were fair.

7. Department periodically discusses and improves peer evaluation practices to maximize utility to instructors and the department.
    a. Rationale: Departmental peer observation practices will affect all faculty in the department who teach. Therefore, some degree of consensus-building is important to help faculty see how peer evaluation helps them meet their own goals, such as being an effective teacher or continuously improving.
    b. Support includes recommendations from one university and our own experiences in the UGA DeLTA project:
        i. University of Oregon: Recommends that departments engage in reflection and conversation about what teaching quality means to them, using this document as a guide. Relatedly, this institution recommends that departments engage in "self-study" about how their peer review processes are serving them using this document as a guide.
        ii. University of Georgia: We observed that departments approached engendering buy-in with faculty in different ways, but all departments that actively worked toward target practices had some discussion among faculty about new teaching evaluation practices. Some departments, especially those that already strongly identified as valuing teaching, found it useful to engage the whole department in specific discussions about practices. Other departments started broader conversations about how teaching was recognized and rewarded, resulting in agreement that evaluation could be better, and then opted to develop and refine new practices with a smaller subset of faculty who were likely to be early supporters.

8. Department provides or arranges formal training about the departmental peer review process for peer observers.
    a. Rationale: Most faculty are not trained as teachers and many do not feel prepared to conduct peer evaluation. Therefore, training can increase comfort and the ability to

productively engage in peer observation. This is aligned with teaching observations used for research purposes, which always suggest a training period

b. Support includes recommendations from multiple universities and indirect support from peer-reviewed research:
    i. University of Oregon: In Peer Review Guide for Departments, mentions a Trained group of observers as a potential way to coordinate the work of peer evaluation
    ii. University of Colorado - Boulder: Peer Evaluation Plans template mentions observers needing to be familiarized with the process before conducting peer observations in their guide for departments - "All first-time observers, or observers who are new to this process, shall meet with the [insert applicable body] to review and become familiar with this process."
    iii. University of Southern California: Online Observation Tool, under the section about the "Characteristics of the Peer Observer," recommends that the Peer Observer "has been trained by CET [Center for Excellence in Teaching] or by a CET Faculty Fellow to use the checklist." They provide a template for a "norming" session for faculty in a department to discuss goals of teaching evaluation, found on the website at the link titled "Facilitating a Norming Session for Peer Review"
    iv. University of Georgia - Departments that piloted new peer observation protocols quickly recognized the need for training sessions to ensure that observers interpreted the peer observation forms in similar ways. This was important even when the observers had collaborated to develop the peer evaluation form.
    v. Research that has aimed to systematically observe teaching always advocates for a training period. Though the standards for reliability among observes are understandably higher for research, this suggests that training is important to help observers pay attention to similar features in a classroom.: All the following papers devote sections to the training of observers in using protocols, emphasizing this as necessary to ensure observers use them in similar ways
        1. Smith et al., 2017 - COPUS
        2. Eddy et al., 2015 - PORTAAL, with separate training manual
        3. Kranzfelder et al., 2019 - CDOP, with supplemental training guide

Reliable: Informed by multiple sources of high-quality evidence

9. Department relies on multiple observations for all instructors, such as using multiple observers, observing multiple lessons, and/or observing multiple courses.
    a. Rationale: One peer observation is just a snapshot of a course and how the instructor teaches that course. The information produced by peer observation is more reliable if it is more expansive than this in at least one way. That could occur in a variety of ways, including getting the perspective of more than one observer, seeing more than one lesson, and/or observing more than one course. Both self-reflection and student voice have the potential to draw on a full semester of experiences, but time constraints make that impossible for peer observation. Establishing a policy for how

multiple observations will be accomplished allows faculty/departments to discuss the data they value most.

b. Support includes recommendations from multiple universities and our own work in the UGA DeLTA project:
    i. University of Colorado – Boulder: leaves this to department on their Peer Evaluation Plans template:
        1. "The second observation can be conducted by the same or different faculty member; this will be decided by the [insert applicable body] in consultation with the observed faculty member and the observer"
        2. BUT also says in the section titled "Procedure for Classroom Observations": "The number of classroom visits is to be determined by the instructor to be observed and the observer. When feasible, 2-3 classroom visits per observation period are recommended. Note that this does not mean filing multiple reports as the final report will be a summation/evaluation of all classroom visitations"
    ii. University of Southern California: Online Observation Tool, in the section about the "Characteristics of the Peer Observer," does NOT ask for multiple observers, but: "Observes a minimum of one class session. Two observations are recommended".
    iii. University of Georgia: Multiple departments determined that observing just one class period did provide a fair assessment of a faculty member's teaching. Two departments opted to rely on at least two observers who attended two different lessons.

10. Department specifies which class materials (e.g., syllabi, exams, homework, slides, handouts) are collected and evaluated as part of peer observation.
    a. Rationale: The learning experiences that students encounter often extend beyond what is observable in the classroom. This includes out-of-class work. Exams and projects also substantially influence students' grades and even their approaches to learning (Stanger-Hall, 2012), but are generally purposely avoided for observation days because they differ from typical instruction. Lastly, the syllabus and other class materials set a tone for the course that can influence student engagement.
    b. Support includes specific recommendations from multiple universities about course materials collected, as well as peer-reviewed research on how class materials provide valuable information about teaching:
        i. University of Colorado - Boulder: Recommendations on the Peer Evaluation Plans template, in the section titled "Procedure for Classroom Observations":
            1. "Observers talk with candidate to properly contextualize observations and review of materials, such as about course goals, lesson goals, class structure, and students"
            2. "Faculty members are required to provide their observer with the course syllabus, and may provide other materials they deem relevant, including access to the learning management system, lesson plans, assessment materials, or outlines explaining the pedagogical goals of classroom activities."

ii. University of Southern California: Part of flowchart that shows how important the sharing of classroom documentation is to the formative process. Also provides a template to review instructional materials from the course (titled "Review of Instructor Assessment Practice Guide") on main page
iii. Vanderbilt University: Offers a list of recommendations for peer evaluation, in which it highlights what evidence of student learning is included, like end-of-course student work, student work throughout the semester, and even student grades
iv. Gin et al., 2021: Argue that syllabi can provide an indication of how inclusive a classroom is by attending to certain sections of the document. Provide a helpful, evidence-based tool to think about inclusion in classroom syllabi (look at Table 2).

11. Department expects observers to talk with instructors to properly contextualize observations and review of materials. This might include discussing course goals, lesson goals, class structure, and students.
    a. Rationale: The learning experiences that students encounter often extend beyond what is observable in the classroom. Additionally, an instructor often has goals or contingencies in mind that are not obvious to an observer. A conversation between the observer and observed offers the chance for a two-way exchange about what influences the observed class period(s), what has preceded and what will follow the observed class period(s). Another feature of these conversations is that they provide the chance for colleagues to talk deeply about teaching, which may be uncommon (Andrews et al., 2016).
    b. Support includes specific recommendations from multiple universities:
        i. University of Colorado - Boulder: Recommendations on the Peer Evaluation Plans template, in the section titled "Procedure for Classroom Observations":
            1. "The observer should talk to the faculty member in advance of the observation in order to understand the objectives for that class and how it fits with the overall course and to provide an overview of the observation to be conducted. During this consultation, it is also recommended that the observer and instructor discuss the possibility of supplementing the course evaluation process with additional forms of data, such as: student interviews"
            2. "The observer should meet with the observed faculty member after the classroom observation(s) but prior to submitting their report to resolve any questions of factual data, discuss concerns or questions the faculty member or observer may have, and to provide formative and constructive feedback to the faculty member. If the evaluation includes multiple classroom visits, you may also schedule meetings in between classroom visits if appropriate"
        ii. University of Southern California: Part of flowchart on formative evaluation practices, which indicates that meeting to "set up logistics and goals" are important to the formative process

iii. University of Arizona: Recommends and provides a rationale for both pre- and post-observation meetings in their Guide for Reviewees and Guide for Reviewers
    1. In the Guide for Reviewees (p. 2), under "Why is a pre-observation meeting important for the peer review of teaching?": "The pre-observation meeting "sets the stage" for a productive observation of your teaching. In this meeting, you and your reviewer will have a chance to discuss your course, including its structure, goals, role in your unit's curriculum, and typical student reactions. You will also be able to discuss the class session to be observed, including your learning goals and planned activities for that day's lesson. This meeting will require about an hour. Before that meeting, please provide your reviewer with a copy of the course syllabus and your written reflection"
    2. In the Guide for Reviewers (p. 1): "Before the pre-observation meeting, your colleague will prepare a written reflection on the target course and send that to you, along with the course syllabus, to read. The goal of the pre-observation meeting is to discuss this written reflection, as well as to prepare for the classroom observation. (If your colleague chose not to do the written reflection in advance of this meeting, please ask her/him to share his verbal responses to some of the questions for reflection.)" And provides questions to help structure conversation.

Longitudinal: Able to document improvement over time & provide feedback

12. Department conducts peer observation over multiple time points in a review period for all instructors to document teaching improvements.
    a. Rationale: A key goal of the DeLTA project is to support continuous teaching improvement. We consider continuous effort to improve a hallmark of effective teaching. In order to recognize and reward improvement in teaching, departments need evidence across time that allows the documentation of change. Therefore, this target practice guides departments to conduct peer observation more than once per review period so that there is the opportunity to document change over time.
    b. Support includes specific recommendations from one university as well as indirect support from peer-reviewed literature on the importance of feedback to teaching efforts:
        i. University of Colorado – Boulder: Articulates clearly that faculty of different ranks have different criteria for the number of evaluations per review period and a rationale for those decisions in their peer evaluation plan template.
        ii. Dillon et al., 2020 (out of University of Oregon): Described the development of a peer observation protocol focused around the types of feedback to offer faculty. All faculty who participated in the pilot implementation reported being more reflective of teaching

13. Department ensures that the peer observation process provides feedback to candidates via follow-up discussion that covers strengths and areas for growth.
    a. Rationale: A key goal of the DeLTA project is to support continuous teaching improvement. We consider continuous effort to improve a hallmark of effective teaching. Feedback is essential to improvement. Our collaborating departments were particularly interested in formative evaluation of teaching that could support the development of faculty as teachers. This target practice focuses on the peer evaluation process providing feedback rather than being strictly evaluative.
    b. Support includes specific recommendations from multiple universities:
        i. University of Colorado - Boulder - From peer evaluation plan template document: "The observer should meet with the observed faculty member after the classroom observation(s) but prior to submitting their report to resolve any questions of factual data, discuss concerns or questions the faculty member or observer may have, and to provide formative and constructive feedback to the faculty member. If the evaluation includes multiple classroom visits, you may also schedule meetings in between classroom visits if appropriate"
        ii. University of Arizona: Recommended in their Guide for Reviewees (with a rationale for both pre & post-observation meetings) and Guide for Reviewers
            1. Guide for Reviewees (p.3) under "Why is a post-observation meeting important for the peer review of teaching?": "As the final step in the peer review of teaching, the post-observation meeting is where your reviewer can share overall impressions of the observed class session and responses on the Classroom Observation Tool, and where the two of you can have a substantive discussion about your course. This meeting will be the most useful if it occurs within a few days of the classroom observation, while the class activities are still fresh in you and your reviewer's minds. This meeting will require about an hour."
            2. Guide for Reviewers (p.2): "The post-observation meeting is most useful if it occurs within a few days of the classroom observation, while the class activities are still fresh in you and your colleague's minds." And provides questions to help structure conversation.
        iii. University of Southern California: Part of flowchart on formative evaluation practices, which indicates that meeting to "debrief" is important to the formative process
        iv. Brickman et al., 2016 found that faculty would value critical feedback from their peers and were more likely to make changes based on peer feedback rather than student feedback if they thought it constructive. However, faculty felt that most peers did not offer critical feedback.

STUDENT VOICE

Structured: Ensures fairness & minimizes bias through formalized processes, collective decision-making, and training

1. Department has formal standards for how and when instructors collect, analyze, and report student data (e.g., response rate expectation, standard quantitative and qualitative analysis).
    a. Rationale: Formal standards set clear expectations for faculty and ensure that the department has sufficient data that is appropriately analyzed to evaluate teaching. In particular, expecting a high response rate increases the representation of student voices captured by these evaluations. Clear expectations for faculty also contribute to equity among faculty through transparency and similar expectations for everyone.
    b. Support comes from recommendations from multiple universities:
        i. University of Colorado - Boulder: Faculty assembly provided clear recommendations for the use of mandatory student evaluations (described at the top of this page) and the university has undertaken pilot studies to establish clear recommendations about how to analyze and use data from mandatory student evaluations

2. Department makes appropriate distinctions in their expectations about student data for different review periods (e.g., annual review, 3rd year review, promotions) and different levels of teaching experience with a given course.
    a. Rationale: This recognizes that different review periods and levels of teaching experience may require different standards of data. Deciding on these distinctions prompts potentially valuable conversations expectations.
    b. Support includes indirect recommendations from a multi-university collaborative report:
        i. British Royal Academy of Engineering's collaborative Career Framework background and overview document (pg. 40): shares ranges of evidence to be used towards evaluation of teaching. Separates between teaching approach and teaching impact, and recommends that early career faculty should focus on collecting evidence of approach, and more experienced faculty focus on evidence of impact.

3. Department periodically discusses and improves expectations for collecting and analyzing data from students to maximize utility to instructors and the department.
    a. Rationale: Departmental student evaluation practices are likely to eventually affect all faculty in the department who teach. Therefore, some degree of consensus-building is important to help faculty see how collecting and analyzing data from students helps them meet their own goals, such as being an effective teacher or continuously improving. Additionally, student data can reflect not only the success of an instructor but the success of a program/department, and therefore may warrant departmental consideration.
    b. Support includes recommendations from one university and our own experiences in the UGA DeLTA project:

i. University of Oregon: Recommends that departments engage in reflection and conversation about what teaching quality means to them, using this document as a guide. Relatedly, this institution recommends that departments engage in "self-study" about how their peer review processes are serving them using this document as a guide.
ii. University of Georgia: We observed that departments approached engendering buy-in with faculty in different ways, but all departments that actively worked toward target practices had some discussion among faculty about new teaching evaluation practices. Some departments, especially those that already strongly identified as valuing teaching, found it useful to engage the whole department in specific discussions about practices. Other departments started broader conversations about how teaching was recognized and rewarded, resulting in agreement that evaluation could be better, and then opted to develop and refine new practices with a smaller subset of faculty who were likely to be early supporters.

4. Department provides or arranges formal training, or other support, for instructors about collecting and analyzing student data, including achieving high response rates, analyzing quantitative and qualitative data systematically and appropriately, gathering data beyond mandatory evaluations, and making comparisons across time.
    a. Rationale: Most faculty are not trained as teachers and many do not feel prepared to collect useful data from students about their learning and course experiences. Training can increase comfort and the ability to productively and efficiently collect and analyze data from students.
    b. Support includes literature that provides strategies for designing, administering and using student evaluation data, and our own experiences in the UGA DeLTA project:
        i. Medina et al., 2019: Recommends that "Course coordinators should receive training with regards to the use of evaluation results as a means to provide feedback to faculty members with the goal of improving course delivery"
        ii. University of Georgia: We observed that faculty and departments lacked experience or historical models of how to achieve high response rates, appropriately analyze mandatory evaluation data, or how to collect and analyze other data about student experiences. Therefore, training for faculty and department leaders will be key developing and successfully enacting more robust departmental practices regarding teaching evaluation using student voices.

Reliable: Informed by multiple sources of meaningful & trustworthy evidence

5. Department expects instructors to do everything they can to achieve high response rates on mandatory student evaluations (e.g., course credit offered, class time set aside).
    a. Rationale: Without a high response rate, the data collected by mandatory student evaluations are not representative of the experiences of students in the class. This may mean that the diversity of student voices are not included, which is unfair to both students and the instructor.

b. Support includes recommendations from multiple universities, particularly from the institutional centers for teaching and learning, suggesting methods to promote response rates:
    i. Iowa State University: Recommends that instructors provide incentives to student to complete mandatory end-of-course evaluations in the Teaching Task Force Report
    ii. Missouri State University Outreach: Seven suggestions to improve evaluation response rate. Excerpts from Evaluating Online Teaching: Implementing Best Practices, by Tobin Mandernach, and Taylor.
    iii. University of California - Berkeley Center for Teaching and Learning: Five suggestions to improve response rate and receive constructive feedback from students (adapted from the University of Oregon website that could not be accessed)
    iv. University of Buffalo Course Evaluations: Specifically states that "administrators should give instructors and students clear, visible examples of how course evaluation data has been used to improve the unit's programs, teaching and facilities." Provides the following as a resource
        1. Berk, R. A. (2012). Top 20 Strategies to Increase the Online Response Rates of Student Rating Scales. *International Journal of Technology in Teaching & Learning, 8*(2).

6. Department recognizes known biases, such as bias against women, minoritized groups, and large class size, and limits comparisons of mandatory student evaluations between instructors.
    a. Rationale: Mandatory student evaluations have been repeatedly demonstrated to be biased based on instructor characteristics that are not relevant to teaching effectiveness, such as gender and race. Class size also influences average ratings. Therefore, it is not valid to compare scores across instructors.
    b. Support includes extensive research demonstrating bias in student evaluations, and recommendations from multiple universities:
        i. Adams et al., 2021: Study found that student's written feedback on evaluations seemed to measure conformity with gendered expectations rather than teaching quality, with particularly negative effects for women.
        ii. Esarey & Valdez, 2020: Study found that even when student evaluations are assumed to be fair, reliable, and without bias, that they can misidentify instructors as skilled and exceptional when they are not, and as poor when they are skilled. Recommends that "course-averaged student evaluation scores should be statistically adjusted to remove any systematic non-instructional influences (i.e. biases) before they are used for any purpose" and that multiple measures are used to evaluate teaching.
        iii. Fan et al., 2019: Study found statistically significant gender and culture bias (against women and those who spoke English with an accent) on student evaluations

iv. Boring, 2015: Study found that male professors receive higher ratings, especially from male students. Models suggest female professors would need to award grades that were 7.5% higher than male professors to achieve similar satisfaction ratings from students according to working paper. Published paper recorded this statistic as actual point differences. (Published paper can be found here: Boring, 2017)

v. Smith & Hawkins, 2011: Study found that Black faculty received lower ratings on student evaluations than white faculty across all averaged items and for global items used in making personnel decisions.

vi. Bedard & Kuhn, 2008: Study found that class size had a negative impact on student evals even when controlling for course and instructor.

vii. Anderson & Smith, 2005: Study found that when instructors behaved according to student expectations (stereotyped gender, ethnicity, and political bias), students rated them better than when instructors did not behave according to expectations. For example, Latina instructors with lenient teaching style were perceived as "warm", but Latina instructors with strict teaching styles rated lower.

viii. Vanderbilt University provides recommendations on using student evaluation with a caution of bias based on older references under the section on Summaries of Research on Student Evaluations

ix. University of Southern California has eliminated mandatory course evaluations entirely. While student evaluations may still be collected, they are not used to evaluate instruction.

x. University of Colorado - Boulder: According to the section on Student Evaluation, the Faculty Assembly recommended that these evaluations be used to provide formative feedback rather than high-stakes assessments: "Many units on campus rely heavily on end-of-semester student evaluations of teaching (SETs, aka FCQs at CU), particularly for annual merit. While students are one of the three key voices in evaluation of teaching, over-reliance on SETs/FCQs as the primary/sole measure of student voice can be problematic (see "Role of Students in Evaluation: Student Voice" in our FAQ). The Boulder Faculty Assembly (BFA) recently recommended that FCQs (SETs) be used primarily as formative feedback rather than summative assessment, that evaluators be made aware of potential bias in FCQs, and that the omnibus questions (rate the instructor and course overall) be removed"

7. Department specifies that quantitative questions on mandatory student evaluations be analyzed as distributions of scores, rather than averages. Because quantitative questions often use an ordinal rating scale (excellent, very good, good, poor), average scores and standard deviations are inappropriate. We cannot assume the points on ordinal scales are equidistant.

a. Rational: Because quantitative questions often use an ordinal rating scale (excellent, very good, good, poor), average scores and standard deviations are inappropriate. We cannot assume the points on ordinal scales are equidistant.
b. Support includes peer-reviewed literature on how to use likert data as well as recommendations from one university based on peer reviewed literature:
    i. Bishop, P. A., & Herron, R. L. (2015). - "Use and Misuse of the Likert Item Responses and Other Ordinal Measures", a review article that summarizes the critiques and supports the use of Likert scale as ordinal vs. interval data. Particularly when there is no middle value, we cannot assume that respondents treat the distance between the options as equal.
    ii. Harpe (2015) - "How to analyze Likert and other rating scale data", offers recommendations for analyzing Likert data. In particular, it stresses not to treat individual items as indicative of continuous data.
    iii. Iowa State University's Center for Excellence in Learning and Teaching - On the Student Evaluation Guidelines in the section under Use, offers evidence-based suggestions on how to read and analyze quantitative data (under the section Additional Information on Use guidelines)

8. Department specifies which set of quantitative student evaluation questions are used for each review period (e.g., annual, promotion).
    a. Rationale: This ensures that the same data are considered for each instructor, avoiding situations where instructors being considered for the same review period are evaluated on different metrics. This may also be important for avoiding situations where instructors opt to only report results from items that reflect most favorably on their teaching.
    b. Support includes a recommendation from a university:
        i. University of Colorado - Boulder: Preliminary report on a pilot project to (re)consider the use of mandatory end-of-course evaluations suggests that departments strategically choose to place emphasis on questions that are important, provide actionable feedback, and tap into students' experiences.

9. Department specifies that student comments on mandatory evaluations be systematically examined to determine teaching strengths and room for improvement.
    a. Rationale: Qualitative student data can provide valuable insights about student experiences. Knowing how to analyse these comments for patterns can help instructors identify successes and challenges in their teaching. It helps instructors and departments avoid the common issues of "cherry picking" highly favorable comments.
    b. Support includes a recommendation from multiple universities:
        i. Worcester Polytechnic Institute page on Interpreting Student Course Reports: Under the section labelled 'Analyzing Student Course Reports Systematically'**:** "Try separating them into piles of high and low ratings of the course or of your teaching. Then within those piles, look for patterns in comments or student demographics that might explain the differences in perspective. The outcome of this analysis should be a few areas or issues for improvement or further exploration."

ii. Lewis, 2002: Offers suggestions for making sense of student comments, accessed from Vanderbilt University's resources on interpreting student evaluations
iii. Iowa State University's Center for Excellence in Learning and Teaching - on this page on the Student Evaluation Guidelines in the section under Analysis and Reporting, offers evidence-based suggestions on how to learn from student comments
iv. University of British Columbia page on Teaching Portfolios: Under the section on Documenting Your Teaching Effectiveness, advocates against selection of student comments, stating that "at UBC, the recommendation is against incorporating select comments from students; ie, you need to include all the comments or none"

10. Department expects instructors to collect, analyze, and interpret some data *beyond* mandatory student evaluations.
    a. Rationale: Given known biases and concerns that mandatory evaluations may not positively correlate with student learning (Braga, Paccagnella, & Pellizzari, 2014; Carrell & West, 2010), those data alone are insufficient to reliably capture student voice in the teaching evaluation process. Measures of student learning and experiences of belonging and inclusion (or lack thereof) in courses may help departments evaluate outcomes they particularly value.
    b. Support includes recommendations from multiple universities:
        i. Many universities recommend mid-semester evaluations, for example:
            1. Vanderbilt University suggests that mid-semester feedback is collected so as to add depth to the narrative
            2. University of Connecticut offers support through their Center for Excellence in Teaching to conduct mid-semester evaluations
            3. Iowa State University's Center for Excellence in Learning and Teaching recommends that student evaluations are used alongside other assessment methods, offering mid-semester feedback as a suggested method under the section on Use of Student Evaluation
            4. Ohio State University offers suggestions for how to gather mid-term feedback
        ii. University of Colorado - Boulder recommends student letters and classroom interviews as other sources of data from students. They provide templates to help elicit valuable feedback from students. Classroom interviews gather feedback from a whole class of students by asking them to reflect on and discuss their learning experiences in the course. These discussions generally occur in the middle of the term so that faculty can receive feedback and take action during the semester. These discussions may be facilitated by a center that supports teaching and learning. The site linked above also provides a template for this process that could potentially be carried out by a colleague.
        iii. University of Oregon developed and approved a midway student experience survey that collects feedback from students that is only available to the instructor and is intended specifically for mid-course adjustments and

clarifying expectations with students. Described here under "Student feedback." Actual questions listed here.

Longitudinal: Able to document improvement over time & provide feedback

11. Department expects instructors to document change (or consistently exemplary results) by comparing data from students across multiple timepoints.
    a. Rationale: A key goal of the DeLTA project is to support continuous teaching improvement. We consider continuous effort to improve a hallmark of effective teaching. Comparing the results of mandatory student evaluations across time points allows faculty to learn and document the ways in which student perceptions of their courses are changing over time. This can provide valuable feedback about specific efforts faculty make to improve. It also emphasizes growth as the main goal, rather than immediate success. Critically, this documentation is also necessary in order to recognize and reward continuous improvement.
    b. Support includes recommendations from multiple universities:
        i. University of Colorado - Boulder: The use of templates to gather student feedback (tailored by and for departmental use) implies that gathering such data over time offers opportunity to compare and document growth of an instructor over time.
        ii. Iowa State University's Center for Excellence in Learning and Teaching: In the Student Evaluation Guidelines, in the section under Instrument (under Additional Information on Instrument guidelines) states that student evaluations are built to provide consistent information over time

SELF VOICE

Structured: Ensures fairness & minimizes bias through formalized processes, collective decision-making, and training

1. Department uses a formal self-reflection form to guide the scope and content of written self-reflection narratives, including standards for what constitutes evidence-based self-reflection. Forms may be adopted or adapted from other departments.
    a. Rationale: A formal self-reflection form helps to standardize the expectations for teaching self-reflection, externalizes what is valued in self-reflection, and can facilitate comparisons across time.
    b. Support includes recommendations from multiple universities, each of which provide faculty with templates to structure their reflection process:
        i. University of Southern California's Templates for Teaching Statements are separated into:
            1. Templates for a three-year evaluation plan, with each template focusing on a different aspect of teaching: Year 1 on active learning strategies; Year 2 on key assignments/assessments; Year 3 on course materials
            2. A template for a single year evaluation plan, which addresses all three areas that are foci for the multi-year plan
        ii. University of Colorado - Boulder's Template for a self-reflective teaching statement for annual review states that "One way of incorporating the voice of the faculty member being evaluated into the evaluation process is to allow space for them to reflect on their own teaching practices in the annual merit process. Typically departments have created a list of guiding questions that align with their values…"
        iii. University of Oregon's template for instructor self-reflection "provides an opportunity for instructors to archive what went well and what might be improved in the future as well as how their teaching aligns with UO's definition of teaching excellence. The reflection also provides a new mechanism for the instructor's own voice to inform evaluators' interpretation of student feedback."
        iv. University of Georgia's own DeLTA Project developed a template to help faculty craft a self-reflection narrative. This walks faculty through the identification of a teaching challenge, the systematic collection and analysis of evidence and the communication of that information for the purposes of formative and summative evaluation. Variations of this template are used by multiple departments.

2. Department periodically discusses and improves standards for written teaching reflections to maximize utility to instructors and the department.
    a. Rationale: Departmental self-evaluation practices are likely to eventually affect all faculty in the department who teach. Therefore, some degree of consensus-building is important to help faculty see how self-evaluation helps them meet their own goals, such as being an effective teacher and continuously improving.

b. Support includes recommendations from one university and our own experiences in the UGA DeLTA project:
    i. University of Oregon: Recommends that departments engage in reflection and conversation about what teaching quality means to them, using this document as a guide. Relatedly, this institution recommends that departments engage in "self-study" about how their peer review processes are serving them using this document as a guide.
    ii. University of Georgia: We observed that departments approached engendering buy-in with faculty in different ways, but all departments that actively worked toward target practices had some discussion among faculty about new teaching evaluation practices. Some departments, especially those that already strongly identified as valuing teaching, found it useful to engage the whole department in specific discussions about practices. Other departments started broader conversations about how teaching was recognized and rewarded, resulting in agreement that evaluation could be better, and then opted to develop and refine new practices with a smaller subset of faculty who were likely to be early supporters. These practices rely, to varying extents, on this document.

3. Department provides or arranges formal training, or other support, for instructors about the self-reflection process and to help instructors meet departmental expectations for documenting self-reflection.
    a. Rationale: Most faculty are not trained as teachers and many do not feel prepared to engage in critical, evidence-based self-reflection about teaching. Therefore, training can increase comfort and the ability to productively engage in teaching self-reflection.
    b. Support includes our own experiences in the UGA DeLTA project:
        i. University of Georgia: Departments that piloted new self-evaluation practices quickly recognized the need for training or discussion sessions to help faculty feel comfortable meeting the new expectations.

Reliable: Informed by evidence and systematic observation

4. Department expects instructors to engage in a self-reflection process, and written documentation thereof, that is focused on tackling teaching challenges (e.g., concerns raised in student evaluations or peer observation, student learning difficulties, lack of engagement).
    a. Rationale: By expecting instructors to focus on teaching challenges, the department communicates that continuous improvement is valued. This encourages faculty to look beyond successes. The process of teaching self-reflection should be relevant to all faculty, as there is always room to improve the teaching in a course.
    b. Support includes recommendations from other universities and our own experiences in the UGA DeLTA project:
        i. University of Southern California: On this page of resources, see "Teaching Statement" and then "Teaching Statement Evaluation Guide." This document is meant to communicate best practices for writing a teaching self-reflection (i.e., teaching statement) for review. Area 5 specifies that self-reflection should "articulate specific areas for improvement or changes to teaching

practices based on student outcomes, USC Student Learning Experience Evaluation data, or other data."

ii. University of Colorado - Boulder template for a self-reflective teaching statement for annual review begins by asking:
    1. "Which aspects of your courses and/or teaching proved to be particularly effective and/or ineffective? How did you assess efficacy?"
    2. "How did your courses go? Please comment on achievement of course goals, level of student engagement, student learning outcomes, and anything else that seems important to you. You may want to address aspects of your teaching that proved to be particularly effective or ineffective. You may focus on one course or several courses."

iii. University of Georgia: Multiple departments have adopted self-reflection practices that ask faculty to document systematic self-reflection on teaching challenges that includes analysis of evidence. These practices rely, to varying extents, on this document.

5. Department expects the self-reflection process, and written documentation thereof, to rely on the systematic analysis of evidence about student learning and experiences.
    a. Rationale: Teaching self-reflection is informative for instructors and for teaching evaluation when it relies on evidence beyond an instructor's own perceptions of what occurs in a course. Critical reflection on additional evidence creates the conditions for continuous teaching improvement because it challenges instructors to make sense of new information about students' experiences.
    b. Support includes recommendations from other universities and our own experiences in the UGA DeLTA project:
        i. University of Southern California: On this page of resources, see "Teaching Statement" and then "Teaching Statement Evaluation Guide." This document is meant to communicate best practices for writing a teaching self-reflection (i.e., teaching statement) for review. Area 3 specifies that a written self-reflection should "point to clear, concise evidence of whether goals were accomplished, and lessons learned" and Area 5 specifies that written self-reflection should "articulate specific areas for improvement or changes to teaching practices based on student outcomes, USC Student Learning Experience Evaluation data, or other data."
        ii. University of Georgia: Multiple departments have adopted self-reflection practices that ask faculty to document systematic self-reflection on teaching challenges that includes analysis of evidence. These practices rely, to varying extents, on this document.

6. Department expectations for self-reflection consider the experience level of instructors. For example, instructors new to a course or teaching undergraduates may primarily rely on informal sources of data (e.g., notes, brief written feedback from students), whereas more experienced instructors rely on formal sources of data (e.g., assessment data) and systematic observation (e.g., feedback from trained peers).

a. Rationale: Faculty at different points in their career will have different levels of teaching experience. This target practice recognizes that the focus and depth of teaching self-reflection should be expected to vary depending on a faculty member's level of teaching experience, experience with a specific course, and experience with critical teaching self-reflection. This practice can guide faculty to collect data appropriate to their level of experience, and ensures departments do not set unreasonable expectations for new faculty..
b. Support includes recommendations from universities and the British Royal Academy of Engineering:
    i. University of Southern California offers templates for a three-year evaluation plan (see Teaching Statements) that asks instructors to focus on a different area of teaching each successive year, implying that expectations change with experience in self-reflection
    ii. British Royal Academy of Engineering's collaborative Career Framework background and overview document, forms of evidence (pg. 40): describes ranges of evidence to be used towards evaluation of teaching. Separates between teaching approach and teaching impact, and recommends that early career faculty should focus on collecting evidence of approach, and more experienced faculty focus on evidence of impact.

Longitudinal: Able to document improvement over time

7. Department expects that written reflections discuss what instructors have achieved and learned since previous self-reflections. This can include learning from both successes and failures that occur as a result of efforts to improve and innovate in the classroom
    a. Rationale: A key goal of the DeLTA project is to support continuous teaching improvement. We consider continuous effort to improve a hallmark of effective teaching. In order to recognize and reward improvement in teaching, departments need evidence across time that allows the documentation of change. This practice emphasizes deliberate reflection on previously outlined goals, and articulation of how those goals were or were not realized. It also recognizes that aiming for continuous improvement will always involve both successes and failures. It normalizes failure and learning from failure as part of the process of improving teaching.
    b. Support includes recommendations and guidelines from universities found on template documents and on institutional websites about reflection:
        i. University websites make clear that self-reflection is a tool for instructors to build upon prior learning, for example:
            1. Ohio State University's page on Documenting Teaching Effectiveness contains a section on Summarizing Your Feedback, which asks instructors to reflect on: "At what point during the quarter do you collect feedback? How often do you collect feedback? For what purposes do you collect feedback? How have you integrated this feedback into your teaching?"

2. Purdue University's Documenting Teaching Effectiveness Overall Plan contains many references to instructors building upon prior evaluations and reflections across the document.
3. University of British Columbia's Teaching Portfolio page: Under the section on Teaching Philosophy recommends that instructors "include the ways that you monitor and evaluate your own teaching and reflect on what the evidence gathered tells you about your teaching". It also asks instructors to reflect "In which ways has your teaching changed in the last five years? Are these changes for the better (for you, for your students)? Explain."

ii. University of Colorado - Boulder template for a self-reflective teaching statement for annual review (link will download a document) asks instructors "How have your efforts and reflections informed changes you have made and/or will make to your teaching practices?"

iii. University of Southern California: Under the Teaching Statement section, the template for Teaching Statement Evaluation Guide requires evaluators to check whether instructors communicate measurable goals from the previous year (#1) and then provide clear evidence as to whether those goals were accomplished and lessons learned (#3)

8. Department expects that written reflections discuss opportunities that instructors have sought to grow and learn as educators, such as mentoring, feedback, collaborations, and teaching professional development.
    a. Rationale: A key goal of the DeLTA project is to support continuous teaching improvement. We consider continuous effort to improve a hallmark of effective teaching. In order to recognize and reward improvement in teaching, departments need evidence across time that allows the documentation of change. This practice places value on efforts to grow and learn, such as participating in teaching professional development. Support includes recommendations from multiple institutions:
    b. Support includes recommendations and guidelines from universities found on template documents and on institutional websites about reflection:
        i. University of Colorado - Boulder's template (link will download a document) for a self-reflective teaching statement for annual review asks instructors to provide information about what they learned from teaching development opportunities (#8) as well as what steps instructors took to develop their pedagogical knowledge and practice (#10b).
        ii. University of Oregon's template for instructor self-reflection (will download a document) asks instructors if they "participate[d] in professional teaching development, and/or engage in campus or National discussions about quality pedagogy and curricula related to this course?" In addition, a following section of their template asks instructors to articulate "how your teaching in this course is informed by research on how students learn and inflected by UO's research mission"

iii. University of British Columbia's Teaching Portfolio page: Under the section on Teaching Activities, dedicates an entire section to the documentation of professional development activities

REFERENCES

Adams, S., Bekker, S., Fan, Y., Gordon, T., Shepherd, L. J., Slavich, E., & Waters, D. (2021). Gender Bias in Student Evaluations of Teaching:'Punish [ing] Those Who Fail To Do Their Gender Right'. *Higher Education*, 1-21. https://doi.org/10.1007/s10734-021-00704-9

Anderson, K. J., & Smith, G. (2005). Students' preconceptions of professors: Benefits and barriers according to ethnicity and gender. *Hispanic Journal of Behavioral Sciences*, *27*(2), 184-201. https://doi.org/10.1177/0739986304273707

Andrews, T. C., Conaway, E. P., Zhao, J., & Dolan, E. L. (2016). Colleagues as change agents: How department networks and opinion leaders influence teaching at a single research university. *CBE—Life Sciences Education*, *15*(2), ar15. https://doi.org/10.1187/cbe.15-08-0170

Baez, B. (2000). Race-related service and faculty of color: Conceptualizing critical agency in academe. *Higher Education*, *39*(3), 363-391.

Bedard, K., & Kuhn, P. (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. *Economics of Education Review*, *27*(3), 253-265. https://doi.org/10.1016/j.econedurev.2006.08.007

Berk, R. A. (2012). Top 20 Strategies to Increase the Online Response Rates of Student Rating Scales. *International Journal of Technology in Teaching & Learning, 8*(2).

Bishop, P. A., & Herron, R. L. (2015). Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *International Journal of Exercise Science*, *8*(3), 297–302.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of public economics*, *145*, 27-41. https://doi.org/10.1016/j.jpubeco.2016.11.006

Braga, M., Paccagnella, M., & Pellizzari, M. (2014). Evaluating students' evaluations of professors. *Economics of Education Review*, *41*, 71-88. https://doi.org/10.1016/j.econedurev.2014.04.002

Brickman, P., Gormally, C., & Martella, A. M. (2016). Making the grade: Using instructional feedback and evaluation to inspire evidence-based teaching. *CBE—Life Sciences Education*, *15*(4), ar75. https://doi.org/10.1187/cbe.15-12-0249

Carrell, S. E., & West, J. E. (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, *118*(3), 409-432. https://doi.org/10.1086/653808

Dillon, H., James, C., Prestholdt, T., Peterson, V., Salomone, S., & Anctil, E. (2020). Development of a formative peer observation protocol for STEM faculty reflection. *Assessment & Evaluation in Higher Education*, *45*(3), 387-400. https://doi.org/10.1080/02602938.2019.1645091

Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE—Life Sciences Education*, *14*(2), ar23. https://doi.org/10.1187/cbe.14-06-0095

Esarey, J., & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assessment & Evaluation in Higher Education*, *45*(8), 1106-1120. https://doi.org/10.1080/02602938.2020.1724875

Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., & Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PloS one*, *14*(2), e0209749. https://doi.org/10.1371/journal.pone.0209749

Gin, L. E., Scott, R. A., Pfeiffer, L. D., Zheng, Y., Cooper, K. M., & Brownell, S. E. (2021). It's in the syllabus… or is it? How biology syllabi can serve as communication tools for creating inclusive classrooms at a large-enrollment research institution. *Advances in Physiology Education*, *45*(2), 224-240. https://doi.org/10.1152/advan.00119.2020

Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about teaching in higher ed: Neglected opportunities to promote change. *CBE—Life Sciences Education*, *13*(2), 187-199. https://doi.org/10.1187/cbe.13-12-0235

Griffin, K. A., & Reddick, R. J. (2011). Surveillance and sacrifice: Gender differences in the mentoring patterns of Black professors at predominantly White research universities. *American Educational Research Journal*, *48*(5), 1032-1057. https://doi.org/10.3102/0002831211405025

Harpe, S. E. (2015). How to analyze Likert and other rating scale data. *Currents in pharmacy teaching and learning*, *7*(6), 836-850. https://doi.org/10.1016/j.cptl.2015.08.001

Kranzfelder, P., Bankers-Fulbright, J. L., García-Ojeda, M. E., Melloy, M., Mohammed, S., & Warfa, A. R. M. (2019). The Classroom Discourse Observation Protocol (CDOP): A quantitative method for characterizing teacher discourse moves in undergraduate STEM learning environments. *PloS one*, *14*(7), e0219019. https://doi.org/10.1371/journal.pone.0219019

Lewis, K. G. (2001). Making sense of student written comments. *New Directions for Teaching and Learning*, *87*, 25-32. https://doi.org/10.1002/tl.25

Medina, M. S., Smith, W. T., Kolluru, S., Sheaffer, E. A., & DiVall, M. (2019). A review of strategies for designing, administering, and using student ratings of instruction. *American Journal of Pharmaceutical Education*, *83*(5). https://doi.org/10.5688/ajpe7177

O'Meara, K., Kuvaeva, A., Nyunt, G., Waugaman, C., & Jackson, R. (2017). Asked more often: Gender differences in faculty workload in research universities and the work interactions that shape them. *American Educational Research Journal*, *54*(6), 1154-1186. https://doi.org/10.3102/0002831217716767

O'Meara, K., Jaeger, A., Misra, J., Lennartz, C., & Kuvaeva, A. (2018). Undoing disparities in faculty workloads: A randomized trial experiment. *PloS one*, *13*(12), e0207316. https://doi.org/10.1371/journal.pone.0207316

Smith, B. P., & Hawkins, B. (2011). Examining student evaluations of Black college faculty: Does race matter?. *The Journal of Negro Education*, 149-162. http://www.jstor.org/stable/41341117

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12*(4), 618-627. https://doi.org/10.1187/cbe.13-08-0154

APPENDIX C: RELEVANT SECTIONS OF INTERVIEW PROTOCOL

Interview protocol:

Did you get a chance to look at the consent form? Any questions?

Is it okay if I start the audio-recorder?

The goal of this interview is to learn about departmental practices related to undergraduate teaching, particularly the practices that may not be part of formal policies and might not be obvious to an outsider. We will start with some questions about teaching evaluation since that is the initial focus of the work of the Leadership Action Team.

Alternate text for non-LAT: The goal of this interview is to learn about departmental practices related to undergraduate teaching, particularly the practices that may not be part of formal policies and might not be obvious to an outsider like me. I'd like to start by asking you about how teaching effectiveness is evaluated in your department.

I'll be asking about a wide array of practices and I don't expect any department to be engaging in all of these practices. So please don't hesitate to say "no" when I ask if something is happening.

1. Can you please talk me through how teaching effectiveness is evaluated for promotion & tenure?
   - Student evaluations?

   - How are peer evaluations used?
   - Is there a particular form that guides this process? OR
   - Are there particular things that you ask observers to pay attention to?
   - Is that written down somewhere?

   - Some departments indicate that colleagues might be very hesitant to provide critical feedback in a peer review for promotion. Do you that happen in your department?

   - Summary of what they said. Is there more I should know about how teaching effectiveness is evaluated from promoted & tenure?

2. To what extent is teaching effectiveness discussed when the department votes on tenure and promotion?
   - What evidence of effectiveness do you think is most strongly valued by your faculty? What makes you think that?

- How does this vary by rank or position type of the candidate?

3. How is the evaluation of teaching effectiveness different for annual review than for promotion & tenure?

- How are student evaluations used?

- IF peer evaluation is mentioned, ask these:
    - Is there a particular form that guides this process? OR
    - Are there particular things that you ask observers to pay attention to?

- Summary of what they have said. Is there more I should know about how teaching effectiveness is evaluated annually?

4. I understand that the university asks you to rate each faculty member with teaching responsibilities as exceeding, meeting, or not meeting expectations for teaching. How do you know when someone exceeds expectations versus meeting expectations?

- If there is a written document, ask if they would be willing to share it.

LAT members only

5. As you know, the initial focus of the LAT is considering how teaching effectiveness is evaluated in our departments. What would you like to see as an outcome for your department in this area?

6. Do you think your departmental colleagues would also value that outcome?

- Alternative text IF they do not have ideas about an outcome they would like to see: Do you think your departmental colleagues see a need for any changes related to how teaching effectiveness is evaluated?