Guidance for Evaluation of Instruction

A Companion to 'Student Perceptions of Learning Experience: Rationale and Broad Principles of Design'

Sub-Committee of the Ad Hoc Committee on Student Perceptions of Teaching Effectiveness

2026-04-25

Table of contents

Preamble

This document was prepared by a sub-committee of the Ad Hoc Committee on Student Perceptions of Teaching Effectiveness as a companion resource. It has not been formally adopted by the full committee and is offered as proposed language for consideration by the Academic Senate Faculty Affairs Committee.

This document provides guidance for the evaluation of instruction in a format that can be directly incorporated into the University Faculty Personnel Policies (UFPP) as §8.3. The section numbering, structure, and language are designed so that, upon approval by the Academic Senate, this text can serve as the basis for the corresponding UFPP section with minimal modification.

This guidance addresses the **summative** evaluation of teaching as understood and required by the UFPP and the Collective Bargaining Agreement — that is, the formal assessment of teaching effectiveness for purposes of retention, tenure, promotion, and other personnel decisions. It is distinct from the **formative** assessment of teaching, which is an informal, voluntary, ongoing process of instructor development offered through the Center for Teaching, Learning and Technology (CTLT). The formative process is described in a separate companion document, Formative Learning Feedback.

§8.3.1. Policy History

[Reserved for Academic Senate.]

§8.3.2. Purpose and Scope

The goal of evaluation of instruction in faculty evaluations is to maintain high quality of instruction and provide guidance to faculty for improvement of instruction. Evaluations of instruction should do the following:

- Briefly and specifically report on the candidate's successes and challenges in instruction.
- Provide brief and specific guidance when important deficits are apparent to evaluators.
- Clearly state any necessary changes to be implemented and documented in the next evaluation cycle.

§8.3.3. Dimensions of Teaching for Evaluation

This guidance adopts the TEval framework (Austin et al., 2025), which identifies seven dimensions of teaching for evaluation. Each dimension encompasses one or more of the criteria currently listed in UFPP §7.2.5.2, as shown below:

Table 1: Mapping of TEval dimensions to current UFPP §7.2.5.2 criteria

TEval Dimension	Current UFPP §7.2.5.2 Criteria
D1: Goals, Content, and Alignment	Competence in the discipline; organization of courses; relevance of instruction to course objectives
D2: Teaching Practices	Ability to communicate ideas effectively; versatility of teaching techniques; appropriateness of teaching techniques
D3: Class Climate	Relationship with students in class
D4: Achievement of Learning Outcomes	Relevance of instruction to course objectives; methods of evaluating student achievement
D5: Reflection and Iterative Growth	Other factors relating to performance as an instructor
D6: Mentoring and Advising	Effectiveness of student advising
D7: Teaching Service, Scholarship, or Community	Other factors relating to performance as an instructor

Proposed revision to UFPP §7.2.5.2

We therefore propose that UFPP §7.2.5.2 be revised to read: "Evaluators shall consider such dimensions as (1) Goals, Content, and Alignment; (2) Teaching Practices; (3) Class Climate; (4) Achievement of Learning Outcomes; (5) Reflection and Iterative Growth; (6) Mentoring and Advising; and (7) Involvement in Teaching Service, Scholarship, or Community."

Evidence for evaluating teaching comes from three sources: the instructor (e.g., CV, syllabi, course materials, student work samples, reflection), peers or observers (e.g., meeting with instructor, class observation, review of student materials), and students (e.g., Student Perceptions of Learning Experience instrument, letters from students). The sections that follow describe each dimension, the guiding questions evaluators should consider, the sources of evidence appropriate to that dimension, and — where applicable — the limitations of particular evidence sources.

§8.3.4. Dimension 1: Goals, Content, and Alignment

This dimension is about what students are expected to learn from the courses taught, whether learning goals are clearly articulated in a way that is accessible to all students, whether course goals are appropriate for the course as part of the larger curriculum and for the audience for which it is intended, whether topics are appropriately challenging and related to current issues in the field, whether the materials are high-quality and aligned with course goals, whether the content represents diverse perspectives, and whether assessments are aligned with course goals.

Sources of evidence: Syllabi, course materials, reflection, meeting with instructor, class observation, review of student materials.

> **Note**
>
> This dimension requires disciplinary expertise to evaluate. It is assessed through peer review (syllabi review, class observation, review of course materials) and instructor self-report (reflection), not through student surveys. The Student Perceptions of Learning Experience instrument does not assess this dimension.

§8.3.5. Dimension 2: Teaching Practices

This dimension is about how in-class and out-of-class time is used, whether assignments, assessments, and learning activities are designed to help all students learn, whether effective or high-impact methods are used to improve understanding and engage all students in learning, whether in- and out-of-class activities provide opportunities for practice and feedback on important skills and concepts, and whether forms of assessment are varied to allow for the success of diverse learners.

Sources of evidence: Syllabi, course materials, reflection, meeting with instructor, class observation, review of student materials.

> **Note**
>
> This dimension requires pedagogical expertise to evaluate. It is assessed through peer observation, review of course materials, and instructor reflection, not through student surveys. The Student Perceptions of Learning Experience instrument does not assess this dimension.

§8.3.6. Dimension 3: Class Climate

This dimension is about the extent to which the class climate reflects regard for students as persons, is supportive, and cooperative, whether it encourages motivation and engagement for all students, whether all students feel included, how student-student and student-instructor dialogue are fostered, what the students' views of their learning experiences are, and how the instructor has sought student feedback and used it to inform their teaching.

Sources of evidence: Syllabi, reflection, class observation, **Student Perceptions of Learning Experience** instrument, letters from students.

This is the only dimension assessed through the **Student Perceptions of Learning Experience** (**SPLE**) instrument, which asks students to report on six aspects of class climate:

- **Regard for Students** — Regard for students as persons in how the instructor interacts with them.
- **Consistent Communication and Enforcement of Expectations** — Equitable treatment and consistent application of standards.
- **Access to Instructor and Instructor Resources** — Perceived accessibility of the instructor for help outside of class.
- **Perceived Course Coherence** — Whether the student could see connections between course elements.
- **Participatory Climate** — Whether the classroom environment supports multiple modes of active engagement.
- **Responsive Learning Environment** — Whether the instructor creates a learning environment that is responsive to all students.

§8.3.6.1. Interpreting Student Perceptions of Learning Experience results

Evaluators and candidates should interpret SPLE results with care, following the scoring, reporting, and visualization guidelines established in the "Student Perceptions of Learning Experience: Rationale and Broad Principles of Design" report. Key principles include:

- **Frequency distributions and percentages, not averages.** SPLE responses are ordered categorical data. They must not be averaged, and evaluators should examine the full distribution of responses, not any summary statistic.

- **No cross-comparisons.** SPLE results must not be compared across instructors, courses, departments, or disciplines. Differences in scores may reflect demographic biases, course characteristics, or nonresponse patterns rather than differences in the learning environment.
- **No extrapolation.** Results from respondents should not be extrapolated to non-respondents. Students who submit evaluations are a self-selected sample of convenience, not a random sample.

§8.3.6.2. Inherent limitations of student evaluation data

Even when student survey items are framed as experiential reports about class climate — as in the SPLE — rather than as evaluative judgments about teaching effectiveness, evaluators must be mindful of the inherent limitations of student evaluation data. These include, but are not limited to, the following factors (Stark, 2026):

- **Gender bias.** Student evaluations have substantial bias from gender: female instructors sometimes receive lower ratings than objectively less effective male instructors; gender affects ratings of ostensibly "objective" items like promptness; and bias varies across disciplines and differs between male and female students (Boring, Ottoboni, and Stark, 2016; MacNell, Driscoll, and Hunt, 2015; Mengel, Sauermann, and Zölitz, 2018).
- **Racial and ethnic bias.** Evaluations show bias from ethnicity and race (Chisadza, Nicholls, and Yitbarek, 2019), and bias against non-native English speakers (Subtirelu, 2015).
- **Age and appearance bias.** Evaluations show bias against older instructors (Bianchini, Lissoni, and Pezzoni, 2013) and bias in favor of physically attractive instructors, especially for female faculty (Wolbring and Riordan, 2016; Babin et al., 2020).
- **Grade expectations.** Evaluations have stronger association with grade *expectations* than with learning (Boring, Ottoboni, and Stark, 2016); students reward grades — not learning — by giving high evaluation scores (Stroebe, 2020).
- **Halo effect.** Students conflate enthusiasm, attractiveness, and other characteristics with effectiveness; enthusiasm is not associated with learning (Feeley, 2002; Keeley et al., 2013; Michela, 2023).
- **Physical environment.** Evaluations are influenced by the physical condition of the room, time of day, mathematical level of the course, class size, and other factors unrelated to instruction (Bedard and Kuhn, 2005).
- **Fabricated responses.** A substantial fraction of students give demonstrably or deliberately false responses (Stanfel, 1995; Clayson and Haley, 2011).
- **Non-random samples.** Response rates are typically below 75%. The respondents are not a random sample; standard statistical measures of uncertainty (standard errors,

confidence intervals) are inapt (Stark, 2026).

- **Perceived learning does not track actual learning.** Students who learn more may report feeling they learned less, and vice versa (Deslauriers et al., 2019; Uttl, White, and Gonzalez, 2017).

§8.3.6.3. Department-associated questions

Departments are not required to add questions to the Student Perceptions of Learning Experience instrument. The university-wide items are designed to provide a comprehensive assessment of class climate across six aspects, and many departments will find them sufficient.

Departments that wish to add questions should weigh the benefit of additional information against the cost of making the instrument more burdensome for students to complete. A longer survey reduces response rates, and lower response rates weaken the representativeness of the data — undermining the very information the additional questions are meant to provide.

If a department elects to add questions, those questions must meet the same standards that govern the university-wide items. The bar is high:

- **Students must be qualified to answer.** The question must concern something students can report on from their own experience, without requiring disciplinary or pedagogical expertise.
- **Students must be able to answer with minimal bias.** The question must elicit an experiential report, not an evaluative judgment. Items that ask students to assess teaching effectiveness, course quality, or instructor competence are not permitted, as these are the items the literature identifies as most susceptible to bias.
- **Closed-ended, structured items only.** Department-associated questions must be closed-ended items on the five-point Likert scale. The university-wide instrument already includes open-ended questions with structured prompts and guardrails designed to minimize equity bias; there is no need for departments to add additional open-ended questions at the department level.

Department-associated questions must use the same five-point ordered categorical (Likert) response scale as the university-wide items (Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree, plus Not Applicable). They must be scored and reported identically to the university-wide questions — as frequency distributions of raw counts and percentages, with no numerical averages, no cross-comparisons, and no extrapolation from respondents to non-respondents. Every guardrail established in the Scoring and Reporting Guidelines of the "Student Perceptions of Learning Experience" report for the university-wide items applies in full force to department-associated questions, lest the protections built into the university-level instrument be undone at the department level.

§8.3.7. Dimension 4: Achievement of Learning Outcomes

This dimension is about whether the instructor clearly communicates the learning goals for the course, what evidence is used to determine the degree to which students achieve the defined course goals, how well course assignments, assessments, and learning activities are aligned with the defined learning goals, whether there are efforts to ensure that all students have equitable opportunities to achieve the learning goals, whether standards for evaluating learning are clear and connected to program, curriculum, or professional expectations, and whether the quality of learning supports success in other contexts.

Sources of evidence: Student work samples, reflection, meeting with instructor, review of student materials.

> **Note**
>
> This dimension is assessed through review of student work, analysis of learning outcomes, and instructor reflection, not through student surveys. The Student Perceptions of Learning Experience instrument does not assess this dimension. While it may seem natural to ask students how much they learned, perceived learning does not track actual learning. In a controlled experiment, Deslauriers et al. (2019) found that students who learned *more* (as measured by test performance) reported feeling they had learned *less*, and vice versa. Uttl, White, and Gonzalez (2017), in a comprehensive meta-analysis, found that the correlation between student evaluation ratings and student learning is effectively zero. A "perceived learning" item would thus measure neither the learning environment nor actual learning, while carrying the same bias vulnerabilities as other evaluative items.

§8.3.8. Dimension 5: Reflection and Iterative Growth

This dimension is about how and why the instructor's teaching has changed over time, whether changes have been informed by evidence of student learning and student feedback, how peer feedback has been incorporated into the instructor's teaching over time, and how the instructor's goals for their courses and students have changed over time.

Sources of evidence: Syllabi, course materials, student work samples, reflection, meeting with instructor.

> **Note**
>
> This dimension is assessed through the candidate's reflective narrative and through longitudinal review of course materials and student outcomes. The Student Perceptions of Learning Experience instrument does not assess this dimension, though SPLE results over time may inform the candidate's reflection.

§8.3.9. Dimension 6: Mentoring and Advising

This dimension is about how effectively the instructor has worked individually with undergraduate or graduate students, whether the instructor establishes clear, individualized, and responsive expectations for student and mentor, whether the instructor provides constructive and timely coaching and feedback, and how the quality of and time commitment to mentoring fit with disciplinary and departmental expectations.

Sources of evidence: CV (student awards, achievements), reflection, meeting with instructor, review of student materials, letters from students.

> **Note**
>
> This dimension is assessed through the candidate's CV, reflective narrative, meeting with the instructor, and letters from students — not through the Student Perceptions of Learning Experience instrument.

§8.3.10. Dimension 7: Involvement in Teaching Service, Scholarship, or Community

This dimension is about how the instructor has contributed to the broader teaching community, both on and off campus. Areas of contribution include the learning culture in the department or institution (e.g., curriculum committees, program assessment, cocurricular activities), engaging with peers on or off campus in teaching communities, workshops, peer reviews, or similar activities, and educational leadership activities (e.g., leading teaching workshops, presentations or publications about teaching, grants related to teaching).

Sources of evidence: CV (participation in teaching and learning committees), reflection, meeting with instructor.

> **Note**
>
> This dimension is assessed through the candidate's CV and reflective narrative. The Student Perceptions of Learning Experience instrument does not assess this dimension.

§8.3.11. Teaching Effectiveness Rubric

The following rubric provides descriptions of teaching practice at three quality tiers — Developing, Proficient, and Expert — for each of the seven dimensions of teaching. This rubric is adapted from the Benchmarks for Teaching Excellence Rubric (University of Kansas Center for Teaching Excellence, 2024).

Evaluators should use this rubric to organize their assessment of each dimension. The rubric is not a checklist; it describes patterns of practice. An instructor may exhibit characteristics of different tiers across different dimensions, and growth across tiers is expected over the course of a career.

Dimension 1: Goals, Content, and Alignment

Developing	Proficient	Expert
Course goals are not articulated, or are unclear, inappropriate, or marginally related to curriculum. Content and materials are outdated or unsuitable. Range of topics is too narrow or too broad. Content is not clearly aligned with curriculum or institutional expectations.	Course goals are articulated and appropriate for curriculum. Content is current and appropriate for topic, students, and curriculum. Course topics have appropriate range. Standard, intellectually sound materials. Course materials reflect multiple viewpoints in the field.	Course goals are well-articulated, high quality, relevant to all students, and clearly connected to program or curricular goals. Content is challenging and innovative or related to current issues in the field. Topics are well-integrated and of appropriate range and depth. High-quality materials, well-aligned with course goals. Course materials reflect multiple perspectives and promote meaningful reflection on them.

Dimension 2: Teaching Practices

Developing	Proficient	Expert
Courses are not sufficiently planned or organized. Practices are not well-executed and show little development over time. Students lack opportunities to practice critical skills. Student engagement is generally low. Assessments are at inappropriate difficulty level or not well-aligned with course goals.	Courses are well-planned and organized. Standard course practices; follows conventions of discipline and institution. Opportunities for practice or feedback on skills embedded in course goals. Practices elicit student engagement. Assessments are appropriately challenging and tied to course goals.	Courses are well-planned and integrated, with meaningful assignments and assessments. Uses effective or innovative methods to support learning in all students. Activities consistently provide opportunities for practice and feedback. Practices foster high levels of active engagement. Assessments are varied and allow students to demonstrate knowledge through multiple modalities.

Dimension 3: Class Climate

Developing	Proficient	Expert
Class climate discourages student motivation or self-efficacy. Does not effectively create a responsive learning environment. Consistently negative student reports of instructor access or interaction. Little attempt to address concerns voiced by students.	Class climate promotes student motivation. Fosters a responsive learning environment with regard for students as persons. No consistently negative student reports of instructor access or interaction. Instructor articulates some lessons learned through student feedback.	Climate promotes motivation, self-efficacy, ownership of learning. Instructor models responsive language and behavior. Fosters an open learning environment that promotes student-student and student-instructor dialogue. Student feedback on instructor access and interaction is generally positive. Instructor seeks and is responsive to student feedback.

For Dimension 3 (Class Climate), the rubric tiers correspond to patterns observable in the SPLE frequency distributions (see the Scoring and Reporting Guidelines of the "Student Perceptions of Learning Experience" report) and in other evidence of class climate.

Dimension 4: Achievement of Learning Outcomes

Developing	Proficient	Expert
Insufficient attention to student understanding; quality of learning is not described or analyzed with clear standards. Evidence of inadequate learning without clear attempts to improve. Quality of learning is insufficient to support success in other contexts.	Standards for evaluating student understanding are clear and generally meet department expectations. Attends to student achievement through formal and informal assessments. Some use of student learning evidence to inform teaching.	Standards for evaluating understanding are clear and connected to program, curriculum, or professional expectations. Consistently attends to student learning, uses it to inform teaching. Efforts to support learning in all students. Quality of learning supports success in other contexts.

Dimension 5: Reflection and Iterative Growth

Developing	Proficient	Expert
Little or no indication of having reflected upon or learned from prior teaching, evidence of student learning, or peer or student feedback. Little or no indication of efforts to develop as a teacher despite evidence of need.	Continued competent teaching, possibly with minor reflection based on input from peers and/or students. Articulates some lessons learned or changes informed by prior teaching, student learning, or feedback.	Regularly adjusts teaching based on reflection on student learning and other feedback, within or across semesters. Examines student performance after adjustments. Reports improved student outcomes based on past teaching modifications.

Dimension 6: Mentoring and Advising

Developing	Proficient	Expert
No indication of effective advising or mentoring (but expected in department).	Some evidence of effective advising and mentoring (define as appropriate for discipline).	Evidence of exceptional quality and time commitment to advising and mentoring (define as appropriate for discipline).

Dimension 7: Involvement in Teaching Service, Scholarship, or Community

Developing	Proficient	Expert
Little or no evidence of positive contributions to teaching and learning culture in department or institution. Little or no interaction with teaching community. Practices and results of teaching are not shared with others.	Some positive contributions to teaching and learning culture in department or institution. Some engagement with peers on teaching. Has shared teaching practices or results with others.	Consistently positive contributions to teaching and learning culture (e.g., curriculum committees, program assessment, co-curricular activities). Regular engagement with peers on teaching. Presentations or publications to share practices or results of teaching with multiple audiences. Scholarly publications or grant applications related to teaching.

§8.3.12. Setting Departmental Expectations

Not all seven dimensions apply to every instructor. Some faculty may play no role in student advising or mentoring; others may not engage in teaching-related service, scholarship, or community activities. The TEval framework recognizes this explicitly: "alternative configurations are possible, and departments, programs, or institutions can customize the dimensions to suit their needs" (Austin et al., 2025, p. 26). Evidence collection should fit each instructor's activities. Departments should identify which dimensions are applicable to each faculty role and evaluate accordingly — an instructor should not be penalized for the absence of activity in a dimension that is not part of their responsibilities.

For the rubric to function as a tool for personnel evaluation, departments and programs must establish and document expectations for each career phase. These expectations should specify the rubric tier expected for each applicable dimension, recognizing that faculty develop across dimensions at different rates and that departmental missions may weight some dimensions more heavily than others. The rubric is intended to guide holistic professional judgment, not to replace it. Evaluators should consider the full pattern of a candidate's teaching practice rather than treating the rubric as a checklist of minimum requirements.

Template for departmental expectations

Departments should adopt language such as the following, adapted to their context and documented in their personnel policies:

"For [*career phase: e.g., retention of tenure-track faculty / tenure / promotion to full professor*]*, the candidate is expected to demonstrate performance at the* [*tier*] *level or above in Dimensions* [*list*]*. A trajectory of growth from* [*tier*] *toward* [*tier*] *is expected in Dimensions* [*list*]*. Performance at the Developing level in any dimension should be accompanied by a documented plan for improvement."*

Specific expectations may vary by department. For example, a department with a strong emphasis on undergraduate mentoring may set higher expectations for Dimension 6 (Mentoring and Advising), while a department with a significant graduate program may weight Dimension 4 (Achievement of Learning Outcomes) more heavily.

Departmental expectations should be:

- Established through faculty-based governance procedures.
- Documented in department or program personnel policies.
- Communicated to candidates in advance of the evaluation cycle.
- Reviewed periodically to ensure alignment with the university's evolving expectations for teaching.

§8.3.13. Training, Resources, and Implementation

The evaluation framework described in this guidance represents a substantial change from current practice. Evaluators should not be expected to implement it without adequate preparation, and instructors should not be expected to navigate it without clear guidance. Prior to implementation, the university must invest in training for evaluators and orientation for instructors.

Training for evaluators. All faculty who serve on peer review committees should receive training on the seven-dimension framework, the teaching effectiveness rubric, and the proper interpretation of Student Perceptions of Learning Experience data — including the inherent limitations described in Chapter . Training should include **norming sessions** in which evaluators from the same peer review committee review sample evidence portfolios and calibrate their application of the rubric. Norming is essential to ensure that evaluators across the university understand and use the instruments and evaluation framework in a coherent way, so that the quality of an instructor's evaluation does not depend on which committee reviews it. Cross-departmental norming sessions are also recommended so that college and university-level review committees apply consistent standards. For this same reason, Deans, and the Provost, should receive the training as well.

Guidance for instructors. Instructors should receive clear guidance on the evaluation framework before their first evaluation cycle under the new system. This guidance should explain the seven dimensions, the rubric tiers, the kinds of evidence that are appropriate for each dimension, and how SPLE data will be used. Instructors should understand what is expected of them at their career phase and how to assemble an evidence portfolio.

Resources and tools. The university should develop and maintain resources to support both evaluators and instructors, including:

- Ready-to-use rubric templates and evidence portfolio checklists.
- A website with guidance documents, sample portfolios, and frequently asked questions — modeled on resources such as the University of Kansas Center for Teaching Excellence and the USC Center for Excellence in Teaching.
- Facilitated workshops for peer review committees at the start of each evaluation cycle.

Phased implementation. To avoid inconsistency — where some evaluations proceed under the new framework while others follow legacy practices, at a cost to instructors — the university should establish a clear implementation timeline with a defined transition date after which all evaluations follow this guidance.