Adopted: June 2, 2026

ACADEMIC SENATE

of

CALIFORNIA POLYTECHNIC STATE UNIVERSITY

San Luis Obispo, CA

AS-1016-26

RESOLUTION ON STUDENT PERCEPTIONS OF LEARNING EXPERIENCE

WHEREAS, In 2013, the Academic Senate adopted AS-759-13, *Resolution on Student Evaluations*, establishing the current university-wide student evaluation questions: (1) "Overall, this instructor was educationally effective" and (2) "Overall, this course was educationally effective"; and

WHEREAS, The value of responses to the two required questions on the student evaluation instruments has come into question as research establishes that items framed as evaluations of teaching effectiveness are susceptible to bias and do not validly measure what they claim to measure (Stark, 2026). This research calls into question the reliability and validity of the responses; and

WHEREAS, The University Faculty Personnel Policies document chapter 8.4 was thus updated and established instrument requirements (8.4.2), Criteria for Conducting Student Evaluation (8.4.3), Procedure for Conducting Student Evaluation (8.4.4), and Student Evaluation Results (8.4.5); and

WHEREAS, Student evaluation instruments as defined in Section 8.4 of the University Faculty Personnel Policies, commonly known as "course evaluations" or "student opinion surveys" shall be referred to throughout this document as "student evaluations of teaching"; and

WHEREAS, Use of online student evaluations commenced in Spring 2015 and has continued since. Student response rates have ranged from a high of 66% in Fall 2016 to a low of 35% in Spring 2020 and have remained in the 40% range since Fall 2021 across the University; and

WHEREAS, Low response rates for the student evaluation instrument compound the existing measurement concerns by introducing non-response bias, further limiting the interpretability of the results; and

WHEREAS, According to the Collective Bargaining Agreement, CBA 2022-2025, “written or electronic student course evaluations of faculty instructional effectiveness... shall be required for all faculty unit employees who teach;” and

WHEREAS “During the time of periodic evaluation and performance review of a faculty unit employee, the Working Personnel Action File (WPAF), which includes all information, materials, recommendations, responses, [course evaluations summaries,] and rebuttals, shall be incorporated by reference into the Personnel Action File;” and

WHEREAS “Faculty unit employees may submit written rebuttals to student course evaluations pursuant to Provision 11.2 when it is believed that additional information is needed or in the case of student bias. Evaluators must review such written rebuttals when reviewing underlying student course evaluations;” and

WHEREAS "Student course evaluations collected as part of the regular student evaluation process shall be anonymous and identified only by course and/or section;” and

WHEREAS “The format of student course evaluations shall be quantitative (e.g., “Scantron” form, etc.) or a combination of quantitative and qualitative (e.g., space provided on the quantitative form for student comments);” and

WHEREAS, In 2025, ASI Resolution #25-04 called for reform of the student evaluation instrument and processes, citing concerns about bias, validity, and the impact of evaluations on faculty from underrepresented groups; and

WHEREAS, The Academic Senate established the Ad Hoc Committee on Student Perception of Teaching Effectiveness, charged with providing a revised policy and resolution to replace AS-759-13; and

WHEREAS, Research on student evaluations of teaching establishes that student evaluations have large unpredictable biases linked to the instructor's gender, race, and other characteristics protected under employment law, and that these biases systematically disadvantage women and faculty from marginalized groups (Stark, 2026); and

WHEREAS, Research has also shown that student evaluations of teaching are subject to measurement bias, in that the results are more likely to have been shaped by discipline, student interest in the course, class level and difficulty, class meeting

time, location, and setting, but not actual instructor or course effectiveness (Boring, Ottoboni and Stark, 2016; Stark and Freishtat, 2014); and

WHEREAS, Research indicates that students are well-positioned to report on their own experience of the learning environment (Stark and Freishtat, 2014; Austin et al., 2025) but are not positioned to assess an instructor's disciplinary competence, pedagogical technique, or the degree to which learning outcomes were achieved (Deslauriers et al., 2019; Uttl, White, and Gonzalez, 2017); and

WHEREAS, How the data collected from student evaluations is interpreted and reported by evaluators is inconsistent across the university — including, in some cases, the use of numerical averages and cross-comparisons that are inappropriate for ordinal categorical data; and

WHEREAS, This makes it difficult, if not impossible, to maintain consistency of interpretation of results from student evaluations within and across colleges; therefore, be it

RESOLVED: That Cal Poly Academic Senate endorses the Ad Hoc Committee's report "Student Perceptions of Learning Experience: Rationale and Broad Principles of Design"; and be it further

RESOLVED: That the Academic Senate establish a standing committee with broad representation charged with oversight of all student feedback instruments and initiatives—summative, formative, and college or department-originated—including future revisions to the instrument; and be it further

RESOLVED: That the Academic Senate forward the companion document "Formative Learning Feedback: A Companion to the Student Perceptions of Learning Experience Report", prepared by a sub-committee of this Ad Hoc Committee, to the Academic Senate Instruction Committee and to the Center for Teaching, Learning and Technology for their consideration; and be it further

RESOLVED: That the Academic Senate forward the companion document "Guidance for Evaluation of Instruction" (formatted as proposed language for UFPP §8.3), prepared by a sub-committee of this Ad Hoc Committee, to the Academic Senate Faculty Affairs Committee for their consideration.

Proposed by: Ad Hoc Committee on Student Perception of Teaching Effectiveness
Date: April 28, 2026

Student Perceptions of Learning Experience

Rationale and Broad Principles of Design

A Report

by

the Ad Hoc Committee on

Student Perceptions of Teaching Effectiveness

2026-05-27

Table of contents

Committee Members

Name	Affiliation
John Pan (Chair)	Industrial and Manufacturing Engineering, CENG
Anelise Sabbag	Statistics, COSAM
Shantanu Kumar	Construction Management, CAED
Laura Cacciamani	Psychology and Child Development, CLA
Eduardo Zambrano	Economics, OCOB
Jermaine Washington	Architecture, CAED (Lecturer)
Carrie A. Langner	Psychology and Child Development, CLA (Department Head)
Jett Palmer	Student Representative (ASI)
Silvia Marijuan	World Languages and Cultures, CLA (CFA)
Kylie Parrotta	Social Sciences, CLA (Faculty Affairs Committee)
James Bingaman	Agricultural Education and Communication, CAFES (Instruction Committee)
Simone Aloisio	Academic Personnel (ex officio, non-voting)
Jean Lee	Academic Programs and Planning (Faculty Director)
Patrick O'Sullivan	Center for Teaching, Learning and Technology (ex officio)

Executive Summary

The Academic Senate charged this committee with providing a revised policy and instrument to replace AS-759-13, which established the current university-wide student evaluation questions in 2013.

The current instrument asks students to assess "educational effectiveness" — a judgment that peer-reviewed research has shown to be susceptible to bias linked to the instructor's gender, race, and other characteristics, and that does not reliably measure teaching quality. Students are, however, uniquely positioned to report on their own experience of the learning environment — whether they felt the instructor engaged with them as individuals, whether expectations were clear and consistently applied, and whether the environment supported their participation and learning.

Building on this distinction, and drawing on the TEval framework for multidimensional evaluation of teaching (Austin et al., 2025), the committee approved the following five motions:

1. **Rename the instrument** to Student Perceptions of Learning Experience (SPLE).
2. **Adopt six aspects of class climate** as the focus of the instrument: Regard for Students, Consistent Communication and Enforcement of Expectations, Access to Instructor and Instructor Resources, Perceived Course Coherence, Participatory Climate, and Responsive Learning Environment.
3. **Retain open-ended questions** tied to specific aspects of class climate through structured prompts, accompanied by an informational anti-bias preamble.
4. **Report Likert-scale results as frequency distributions** — raw counts together with percentages — excluding the use of means and medians, and provide guidance for the proper interpretation of these results.
5. **Administer the instrument using a hybrid approach** where online surveys are completed during in-class time in the last two weeks of instruction before finals.

The remainder of this report provides the rationale, evidence base, and implementation details for these recommendations.

Chapter 1

Introduction

1.1 Background

In 2013, the Academic Senate adopted AS-759-13, establishing the current university-wide student evaluation questions. In 2025, ASI Resolution #25-04 called for reform of the evaluation instrument and processes. In turn, the Academic Senate established the Ad Hoc Committee on Student Perception of Teaching Effectiveness, charged with providing a revised policy and resolution to replace AS-759-13.

The committee was given the following charges:

1. Reviewing the reliability and validity of the prompts required on all student evaluation instruments, suggesting revisions to the prompts if necessary, and determining if additional prompts are necessary to obtain a more reliable and valid assessment of teaching effectiveness at Cal Poly.
2. Revisiting the criteria for procedures for conducting student evaluations to increase response rates and reduce incidences of bias, particularly negative bias toward people of color, women, and other minoritized populations in student feedback.
3. Reviewing how both quantitative and qualitative data collected from student evaluations are provided to faculty, the analysis of the data, and how data are presented for review for retention, promotion, and tenure. The committee should also consider as part of their charges what data is appropriate for development of teaching effectiveness purposes and data appropriate for performance evaluation.
4. Suggesting processes for disseminating results of student evaluations to Cal Poly students.[1]

[1] After consulting with the Academic Senate Faculty Affairs Committee, Academic Personnel, and CFA, the committee learned that it is not possible to share course evaluation survey information with students under the current CBA and therefore did not pursue this charge further.

1.2 Summary of recommendations

The proposal this committee has crafted has five parts.

First, it unanimously recommends that the instrument known as the **Student Evaluation of Instruction** and **Student Evaluation of Faculty** in the University Faculty Personnel Policies[2] be renamed as **Student Perceptions of Learning Experience**.

Second, it unanimously recommends for the aspects of teaching effectiveness assessed through the **Student Perceptions of Learning Experience** instrument to be the following:

Interpersonal — how the instructor relates to individual students:

1. Regard for Students
2. Consistent Communication and Enforcement of Expectations
3. Access to Instructor and Instructor Resources

Structural — how the course is experienced as a whole by the students:

4. Perceived Course Coherence

Environmental — what the classroom feels like as a shared space:

5. Participatory Climate
6. Responsive Learning Environment

In the context of this recommendation, the committee unanimously recommends that the Academic Senate establish a standing committee with broad representation charged with oversight of all student feedback instruments and initiatives—summative, formative, and department-originated—including future revisions to the instrument.

Third, the committee discussed whether to remove open-ended questions from the summative instrument due to the extensive evidence of bias in unstructured student comments (see Evidence on bias in open-ended comments below). A motion to remove them from the Student Perceptions of Learning Experience obtained three votes in favor and five votes against. The motion failed. The committee then voted unanimously to retain open-ended questions in the Student Perceptions of Learning Experience instrument under structured prompts and guardrails designed to minimize bias (see Guardrails for open-ended questions below).

Open-ended questions remain a key component of the companion Formative Learning Feedback proposal, where they serve their intended developmental purpose.

The committee understands that the final decision on open-ended questions rests with the Academic Senate, and that endorsement of this report does not commit the Senate to either approach.

[2] In sections 3.2, 3.4, 7.2, 8.1 and 8.4.

Fourth, it unanimously recommends that Likert-scale results be reported as frequency distributions — raw counts together with percentages — excluding the use of means and medians, with guidance for the proper interpretation of these results (see Chapter 4).

Fifth, it unanimously recommends a hybrid approach where online surveys are completed during in-class time in the last two weeks of instruction before finals (see Chapter 5).

This report is organized as follows. Chapters 2 and 3 provide the rationale for the instrument's name and design. Chapter 4 establishes scoring and reporting guidelines. Chapter 5 addresses implementation best practices. The appendix presents a sample survey instrument with a recommended preamble and sample items. The items presented in the sample survey are illustrative. They are intended to demonstrate how the six aspects of class climate can be operationalized as experiential survey items. **The sample survey is not intended to be the final instrument.**

Chapter 2

Student Perceptions of Learning Experience (SPLE)

The name change recommendation in this chapter was approved unanimously by the committee.

2.1 Rationale for the name change

The current names — **Student Evaluation of Instruction** and **Student Evaluation of Faculty** — mischaracterize what the instrument does and should do. The word "evaluation" implies that students are rendering a verdict on the quality of instruction or on the instructor. They are not. As detailed below, the proposed instrument asks students to report on their own experiences in the classroom: whether they felt treated with regard, held to consistent standards, able to access help, able to see how course elements connected, comfortable participating, and that the learning environment was responsive to them. These are experiential reports, not evaluative judgments.

This distinction is not merely semantic. The peer-reviewed literature on student evaluations of teaching (SET) establishes that items framed as evaluations of teaching effectiveness, course effectiveness, or instructor competence are particularly susceptible to bias — including bias linked to the instructor's gender, race, and accent — and are evidently misleading (Boring, Ottoboni, and Stark, 2016; Stark, 2016; Stark, 2026). By contrast, items that ask students to report on their own experience are less susceptible to these biases, precisely because they do not ask students to make judgments they are not qualified to make. The name of the instrument should reflect what it actually measures.

2.1.1 Sources of evidence about the validity and bias of SET

Some earlier research has argued that student evaluations are valid and reliable measures of teaching effectiveness (e.g., Marsh, 1987; Abrami, 2001; Berk, 2005). This committee exam-

ined this claim in light of the more recent experimental and quasi-experimental evidence summarized below.

Studies that claim SET are fair and valid rely on data that cannot answer the relevant question. Some studies compare average SET for male and female faculty and conclude there is no bias because these averages are similar. That conclusion is unwarranted because "one cannot assess gender bias in SET merely by comparing how women and men are rated by students: that comparison does not control for actual differences in teaching effectiveness, subject matter, class size, format, etc., resulting in confounding (Boring et al., 2016; Wagner et al., 2016). The appropriate question is not 'do men and women get similar ratings?' but rather 'would a given instructor teaching a given course have received different ratings if their gender had been different but nothing about their teaching were different?' " (Stark, 2026, p.7).

Randomized experiments and natural experiments — where nature assigns subjects to treatments as if at random — in real class settings provide the strongest evidence about whether SET measure teaching quality or something else. Such research has found:

- SET have weak or negative association with objective measures of learning (Carrell and West, 2010; Braga et al., 2014; Boring et al., 2016)
- SET have substantial bias from gender: female instructors sometimes get lower ratings than objectively less effective male instructors (Boring et al., 2016); gender affects ratings of "objective" items like promptness (MacNell et al., 2015; Boring et al., 2016); bias varies across disciplines (Boring et al., 2016; Mengel et al., 2018); the bias of male and female students towards male and female faculty differs (Boring et al., 2016)
- SET have bias from ethnicity and gender (Chisadza et al., 2019)
- SET have stronger association with grade expectations than with learning (Boring et al., 2016)
- Students reward grades — not learning — by giving high SET scores (Cho et al., 2015; Carrell and West, 2010; Braga et al., 2014; Stroebe, 2020)
- Providing cookies during class increases ratings of instructors and course materials (Hessler et al., 2018)
- The number of points on Likert scales affects gender differences in SET scores (Rivera and Tilcsik, 2019)
- Student perceptions of their learning do not match objectively measured learning (Deslauriers et al., 2019; Dunning et al., 2004; Hartwig and Dunlosky, 2017; Knof et al., 2024; Kruger and Dunning, 1999; Lake, 2001; Lindsey and Nagel, 2015; Woolliscroft et al., 1993; Xu et al., 2024)

Source: Stark, 2026, pp. 2–3

Moreover, such research has found that "bias may be large in some situations and small in others... Indeed, the main reason it is impossible to adjust SET for bias is that there are many

sources of bias that may interact in complex ways. SET cannot be presumed to be valid, reliable, or fair in any given course, department, or university, absent affirmative evidence of reliability, validity, and unbiasedness in that time and place." (Stark, 2026, p.8).

It is on the strength of the experimental and quasi-experimental evidence — which can control for these confounds — that this proposal reframes the instrument around experiential reports rather than evaluative judgments, since these are the ones that this literature finds most susceptible to bias.

2.2 The proposed name

Each word in the proposed name — **Student Perceptions of Learning Experience** — is chosen deliberately:

- **Student**: the respondent.
- **Perceptions of**: what the data represent. The word "perceptions" acknowledges that the instrument captures how students experience the learning environment from their own vantage point. Their nature as perceptions is already captured in the Collective Bargaining Agreement (CBA §15.15). Students occupy a position in the classroom that no other observer shares — they are the only ones who can report on whether the instructor engaged with them as individuals, whether they could see how the course fit together, or whether they felt comfortable participating. "Perceptions" names this unique epistemic contribution directly: the data are the students' own account of their experience, grounded in what they are distinctively positioned to observe.
- **Learning Experience**: what is being reported on. "Learning experience" scopes the instrument to the educational context without making the teaching or the instructor the object of assessment. It signals that the data concern the student's experience of learning — the process, not the outcome — rather than a judgment of instructional quality.

Chapter 3

SPLE Questionnaire Design

The recommendations in this chapter — the adoption of the six aspects of class climate as the focus of the instrument, and the retention of open-ended questions under structured prompts with an informational anti-bias preamble — were approved unanimously by the committee.

3.1 Rationale for the aspects of teaching effectiveness chosen to be included in the survey

When determining which aspects of teaching effectiveness should be included in the **Student Perceptions of Learning Experience**, the committee used the following three criteria. This approach is consistent with the broader movement toward multidimensional evaluation of teaching, which recognizes that student surveys should focus on dimensions students are qualified to assess, as part of a comprehensive evaluation system (TEval Project, 2025; Austin et al., 2025).

1. **It carries a summative component.** The dimension is relevant to personnel decisions under the UFPP.
2. **Students are qualified to assess it.** Reporting on the dimension does not require disciplinary or pedagogical expertise (Palmer, 2026; Stark, 2016).
3. **Students can assess it with minimal bias.** The dimension concerns experiential reports rather than evaluative judgments that the literature identifies as particularly susceptible to bias.

A useful starting point for applying these criteria is the TEval framework developed by Austin et al. (2025), an NSF-funded initiative that draws on twenty-five years of scholarly work on teaching evaluation. The framework identifies seven dimensions of teaching for evaluation, each accompanied by guiding questions that articulate what the dimension captures. Together, the seven dimensions provide a comprehensive definition of high-quality educational practice.

The seven dimensions of the TEval framework (Austin et al., 2025)

Guiding questions for each dimension of the framework

Dimension 1: Goals, Content, and Alignment. What are students expected to learn from the courses taught? Are learning goals clearly articulated in a way that is accessible to all students? Are course goals appropriate for the course as part of the larger curriculum and for the audience for which it is intended? Are topics appropriately challenging and related to current issues in the field? Are the materials high-quality and aligned with course goals? Does the content represent diverse perspectives? Are assessments aligned with course goals?

Dimension 2: Teaching Practices. How is in-class and out-of-class time used? Are assignments, assessments, and learning activities designed to help all students learn? What effective or high-impact methods are used to improve understanding and engage all students in learning? Do in- and out-of-class activities provide opportunities for practice and feedback on important skills and concepts? Are forms of assessment varied to allow for the success of diverse learners?

Dimension 3: Class Climate. To what extent is the class climate respectful, supportive, and cooperative? Does it encourage motivation and engagement for all students? Do all students feel included? How are student-student and student-instructor dialogue fostered? What are the students' views of their learning experiences? How has the instructor sought student feedback, and how has feedback informed the instructor's teaching?

Dimension 4: Achievement of Learning Outcomes. Does the instructor clearly communicate the learning goals for the course? What evidence is used to determine the degree to which students achieve the defined course goals? How well are course assignments, assessments, and learning activities aligned with the defined learning goals? Are there efforts to ensure that all students have equitable opportunities to achieve the learning goals? Are standards for evaluating learning clear and connected to program, curriculum, or professional expectations? Does the quality of learning support success in other contexts?

Dimension 5: Reflection and Iterative Growth. How and why has the instructor's teaching changed over time? How have changes been informed by evidence of student learning and student feedback? How has peer feedback been incorporated as changes in the instructor's teaching over time? How have the instructor's goals for their courses and students changed over time?

Dimension 6: Mentoring and Advising. How effectively has the instructor worked individually with undergraduate or graduate students? Does the instructor establish clear, individualized, and responsive expectations for student and mentor? Does the instructor provide constructive and timely coaching and feedback? How does the quality of and time commitment to mentoring fit with disciplinary and departmental expectations?

> **Dimension 7: Involvement in Teaching Service, Scholarship, or Community.** How has the instructor contributed to the broader teaching community, both on and off campus? Areas of contribution can 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; educational leadership activities (e.g., leading teaching workshops, presentations or publications about teaching, grants related to teaching).

When the seven dimensions are assessed against the three criteria above, one dimension stands out as the natural focus of the student survey: **Dimension 3 — Class Climate**. Class climate carries a summative component: the UFPP requires evidence of the instructor's effectiveness in creating a productive learning environment, and how students experience the classroom is directly relevant to that requirement. Students are qualified to assess it: reporting on whether the classroom felt supportive, responsive, and conducive to their learning does not require disciplinary or pedagogical expertise — it requires only that students reflect on their own experience. And students can assess it with minimal bias: items about class climate elicit experiential reports ("I felt treated with regard," "I felt the instructor created a learning environment that was responsive to all students") rather than the evaluative judgments about teaching effectiveness or instructor competence that the literature identifies as particularly susceptible to bias.

This focus on class climate is particularly fitting at a polytechnic university organized around Learn by Doing. Where students learn primarily through active engagement — in labs, studios, clinics, and collaborative projects — the learning environment is not a backdrop to instruction but the literal space in which learning happens. The climate of that space is, accordingly, not a secondary concern but a direct determinant of whether the pedagogy works.

The remaining six dimensions, by contrast, do not meet all three criteria. Dimensions 1, 2, and 4 — concerning course goals, teaching methods, and achievement of learning outcomes — require disciplinary or pedagogical expertise that students do not possess, and items targeting these dimensions are among those most susceptible to bias (Stark, 2016; Boring, Ottoboni, and Stark, 2016; Stark, 2026). Dimensions 5, 6, and 7 — reflection and growth, mentoring, and service — concern activities that students in a single course generally cannot observe or are not positioned to evaluate. It is worth noting that items asking students whether they feel they learned a great deal — while intuitively appealing — fall squarely within Dimension 4. The peer-reviewed evidence shows that 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 — a strong anti-correlation between perceived and actual learning. Uttl, White, and Gonzalez (2017), in a comprehensive meta-analysis correcting for small-sample and publication bias, found that the correlation between student evaluation ratings and

student learning is effectively zero. As Stark (2026) summarizes, student perceptions of their learning do not match objectively measured learning — a finding replicated across multiple disciplines and study designs. A "perceived learning" item would thus measure neither the learning environment nor actual learning, while carrying the same bias vulnerabilities as other evaluative items.

Having identified class climate as the appropriate focus, the committee then asked: *how can class climate be assessed comprehensively, with aspects that are conceptually distinct and collectively exhaustive?* The guiding questions for Dimension 3 in the TEval framework point toward the answer. They ask whether the climate reflects *regard for students as persons*, is *supportive*, and *cooperative*; whether it encourages *motivation and engagement*; whether all students feel *included*; how *dialogue* is fostered; and what students' *views of their learning experiences* are. Drawing on these guiding questions — and on the broader literature on classroom climate (Moos, 1979; Fraser, Treagust, and Dennis, 1986; Fraser, 1998; Lizzio, Wilson, and Simons, 2002; Frisby and Martin, 2010; Ambrose et al., 2010; Hurtado et al., 2012; Hagenauer and Volet, 2014) — the committee identified six aspects, each capturing a distinct facet of the student's experience in the classroom. In arriving at these aspects, the committee reviewed student course evaluation survey questions currently used by colleges and departments at both the San Luis Obispo and Solano campuses. Furthermore, in naming these aspects, the committee was deliberate in selecting language that describes what students experience without invoking terms that the literature associates with gendered or racialized expectations in evaluation contexts. These are described below.

Interpersonal — how the instructor relates to individual students

3.1.1 Regard for Students

What it captures: Whether the instructor engages with students as individuals — acknowledging their contributions, responding to their questions with care, and treating them as persons whose presence and participation matter.

> How Regard for Students differs from the other aspects
>
> An instructor can apply the same standards to everyone yet be dismissive in manner. A class can welcome questions without the instructor showing regard for the students offering them. A course can feel *coherent* (well-structured, connected) while the instructor is curt or condescending. Regard for Students is about the quality of interpersonal treatment, not consistency of standards (Consistent Communication and Enforcement of Expectations), availability outside class (Access to Instructor and Instructor Resources), perceived course structure (Perceived Course Coherence), conditions for engagement (Participatory Climate), or belonging (Responsive Learning Environment).

3.1.2 Consistent Communication and Enforcement of Expectations

What it captures: Whether expectations are communicated clearly and applied consistently — no favoritism, uniform access to learning and assessment.

> How Consistent Communication and Enforcement of Expectations differs from the other aspects
>
> An instructor can show *regard for students as persons* while playing favorites. A class can feel responsive in atmosphere while grading or attention is unevenly distributed. A course can be *coherent* (activities clearly connected to goals) while standards are applied inconsistently. Consistent Communication and Enforcement of Expectations is about equity across students, not the character of interaction (Regard for Students), availability outside class (Access to Instructor and Instructor Resources), perceived course structure (Perceived Course Coherence), the openness of the environment (Participatory Climate), or sense of belonging (Responsive Learning Environment).

3.1.3 Access to Instructor and Instructor Resources

What it captures: Whether the student can access the instructor and the resources the instructor provides — office hours, email, after-class conversations, course materials, and other support for learning.

> How Access to Instructor and Instructor Resources differs from the other aspects
>
> An instructor can be available one-on-one but create a poor in-class climate (Participatory Climate). A student may find the instructor easy to reach but, once there, feel dismissed (Regard for Students) or experience uneven standards (Consistent Communication and Enforcement of Expectations). A course can be *coherent* in structure while the instructor is difficult to reach outside of class. Access to Instructor and Instructor Resources is about availability, not the quality of what happens during interaction (Regard for Students), consistency of standards (Consistent Communication and Enforcement of Expectations), perceived course structure (Perceived Course Coherence), in-class environment (Participatory Climate), or belonging (Responsive Learning Environment).

Structural — how the course is experienced as a whole

3.1.4 Perceived Course Coherence

What it captures: Whether the student could see connections between course elements — that what happened in class, what was assigned, and what was assessed were recognizably related. A course can be highly coherent — readings connect to lectures connect to assessments — even when the content is disorienting or challenges students' prior beliefs.

Perceived Course Coherence might appear to belong with course design (Goals, Content, and Alignment) rather than with class climate. But what a peer reviewer assesses from the syllabus — whether the course elements are aligned — is different from what the student experiences in the classroom — whether the connections between those elements are *visible* to them. A syllabus can be perfectly aligned on paper while students experience the course as disjointed because the connections were never made explicit. It is this experiential dimension — perceived structure, not designed structure — that the SPLE measures, and that makes Perceived Course Coherence a class climate variable.

> How Perceived Course Coherence differs from the other aspects
>
> A class can score well on every other aspect — students treated with regard, standards applied consistently, instructor available, environment participatory and responsive — while the student still cannot see how the pieces fit together, how today's class connects to last week's, or how the assessments relate to what was covered. Perceived Course Coherence captures one specific, concrete experience: whether the student could see the connections between course elements. It is not a summary of the overall learning experience or a proxy for teaching effectiveness. It is about perceived structure, not interpersonal treatment (Regard for Students), consistency of standards (Consistent Communication and Enforcement of Expectations), availability outside class (Access to Instructor and Instructor Resources), conditions for engagement (Participatory Climate), or belonging (Responsive Learning Environment).

Environmental — what the classroom feels like as a shared space

3.1.5 Participatory Climate

What it captures: Whether the classroom environment supports multiple modes of active engagement — asking questions, sharing ideas and/or resources, discussing with peers, and making mistakes without penalty. This aspect concerns the conditions for engagement, not the format of instruction — a lecture in which the instructor welcomes questions and responds to them thoughtfully is a participatory climate no less than a seminar built around discussion.

> How Participatory Climate differs from the other aspects
>
> A class can be participatory in structure while individual students still don't feel they *belong* (Responsive Learning Environment). The instructor can show regard for students in replies without the environment actually encouraging participation. A course can feel *coherent* (well-structured) while the classroom format discourages questions, discussion, or student-to-student dialogue. Participatory Climate is about the conditions for engagement in class — including peer interaction — not the quality of treatment (Regard for Students), consistency of standards (Consistent Communication

and Enforcement of Expectations), availability outside class (Access to Instructor and Instructor Resources), perceived course structure (Perceived Course Coherence), or belonging (Responsive Learning Environment).

3.1.6 Responsive Learning Environment

What it captures: Whether the instructor creates a learning environment that is responsive to the range of students in the class — one that reflects awareness of differences in background, preparation, learning needs, and experience, rather than treating all students as interchangeable.

How Responsive Learning Environment differs from the other aspects

A student can be treated with regard and held to consistent standards without feeling they belong. A class can be *participatory* (questions encouraged, ideas welcomed) while a student still feels like an outsider — because of whose experiences are centered, who dominates discussion, or what the implicit culture of the class signals. A course can be *coherent* (activities connect, expectations are clear) while a student feels the class was not designed with them in mind. Responsive Learning Environment is about belonging in the group, not individual treatment (Regard for Students, Consistent Communication and Enforcement of Expectations), one-on-one availability (Access to Instructor and Instructor Resources), perceived course structure (Perceived Course Coherence), or conditions for engagement (Participatory Climate).

3.2 Evidence on bias in open-ended comments

The committee reviewed the following evidence on bias in open-ended comments. This evidence informed the committee's unanimous decision to retain open-ended questions only under the structured prompts and guardrails described in the next section.

The design of the **Student Perceptions of Learning Experience** rests on a principle: ask students only about things they are qualified to report on, in a form that minimizes bias. The Likert-scale items above are carefully worded to elicit experiential reports — structured statements about what the student felt — rather than open-ended evaluative judgments. An unstructured open-ended question undoes this by design.

Open-ended comments reintroduce exactly the biases the instrument is built to exclude

When given an unstructured prompt, students are free to comment on anything — teaching effectiveness, grading leniency, course organization, the instructor's appearance, accent, or personality — all topics the literature identifies as particularly suscepti-

ble to bias (Boring, Ottoboni, and Stark, 2016; Stark, 2026). The structured Likert items constrain responses to experiential reports about class climate; an open-ended field removes that constraint entirely.

⚠ The research on open-ended comments is clear

In a controlled experiment where identical online courses were taught under male and female instructor names, students commented on women's appearance and personality far more often than men's (Mitchell and Martin, 2018). An analysis of over 14 million reviews found that male professors were described as "brilliant" or "genius" two to three times more often than female professors across every field studied (Storage et al., 2016). A survey of 674 academics found that the highest volume, most derogatory, and most threatening abuse in student evaluations is directed at women and academics from marginalized groups — leading the authors to conclude that anonymous comments in student evaluations must be removed if institutions wish to be inclusive (Heffernan, 2023). A review of over 100 articles on SET bias concluded that "the clearest evidence of gender bias is in qualitative comments" and recommended that institutions "restrict or eliminate the use of qualitative comments" (Kreitzer and Sweet-Cushman, 2021). The harm to academics is considerable: a survey of 741 Australian academics who had received anonymous non-constructive student commentary found significant adverse effects on mental health, stress, and professional confidence, with younger and tenured academics reporting the greatest vulnerability (Hutchinson et al., 2024). At one large university, a machine-learning screening system flagged 6.9% of all student comments — 4,258 out of 62,049 — as potentially harmful; manual screening at that scale is not feasible, which means institutions that include open-ended comments in the personnel file are including content they cannot even review (Gibson et al., 2022).

⚠ Open-ended comments resist the reporting standards this proposal establishes

The scoring and reporting methodology for the **Student Perceptions of Learning Experience** that this Committee recommends — frequency distributions, no numerical averages, no cross-comparisons — is designed to prevent misinterpretation and misuse of the data. Open-ended comments cannot be reported as frequency distributions, cannot be standardized, and invite selective quotation by evaluators. A single vivid comment, whether positive or negative, can disproportionately influence a reader in ways that a frequency distribution of structured responses does not (Boysen et al., 2014; Linse, 2017).

A growing number of institutions are restricting or removing open-ended comments

Cal Poly Pomona, the University of Houston, and Florida State University restrict open-ended comments so that only the instructor can see them. St. Olaf College's Office of Institutional Effectiveness states plainly: "Invitations for open-ended comments should be avoided, as these tend to produce the strongest evidence of bias." USC eliminated student evaluations from tenure and promotion decisions in 2018. UCLA made them optional for personnel actions in 2024. Miami University's policy states that evaluations "will be conducted for formative purposes only." The University of Toronto Faculty Association, as of January 2026, has an active grievance at arbitration challenging the use of student evaluations, citing discriminatory, harassing, and abusive comments that members receive in the open-ended portions of their evaluations.

This does not mean students should have no voice beyond the six items. It means that unstructured feedback belongs in the formative component of the evaluation of teaching — a separate, developmental process designed exclusively to help the instructor grow as an educator (Centra, 1993; Berk, 2005). Best practices are that formative results are shared only with the instructor and are not used for employment decisions (Benton and Young, 2018; Stark and Freishtat, 2014). In this context, open-ended questions can serve their intended purpose without the risk of biased comments influencing personnel decisions.

3.3 Guardrails for open-ended questions

If the Academic Senate elects to retain open-ended questions in the Student Perceptions of Learning Experience, which is this committee's recommendation, the committee recommends the following guardrails:

The committee recognizes that open-ended questions provide qualitative information that structured items alone cannot capture — including the ability to surface concerns the instrument designers did not anticipate and to give students a voice in their own words. In a listening session with ASI on April 13, 2026, students expressed support for retaining open-ended questions, noting that they allow students to provide context for their Likert-scale responses and to offer suggestions for improvement. Their input helped shape the committee's decision to retain open-ended questions under structured prompts. The TEval framework emphasizes that effective evaluation involves "multiple lenses," and that the student lens captures experiences invisible to peer reviewers and self-reports (Austin et al., 2025). It is for this reason that the committee voted to retain open-ended questions rather than remove them, while adopting the guardrails described below to mitigate the equity bias that the literature documents in unstructured responses.

1. **Informational framing.** The preamble must explicitly instruct students to comment on specific aspects of their learning experience and to avoid comments about the instructor's personal characteristics, consistent with the anti-bias framing described in

Section 5.5 (Boring and Philippe, 2021).

2. **Structured prompts, not generic invitations.** Open-ended questions must not use generic prompts such as "Please comment on the instructor" or "What are your suggestions for improvement?" Instead, each open-ended question should be tied to a specific aspect of class climate — for example, asking the student to elaborate on their experience of class coherence. This channels comments toward the dimensions the instrument measures and away from the unstructured commentary that the literature identifies as most susceptible to bias.

Residual bias in open-ended responses. The evidence on informational framing and structured prompts is encouraging for structured Likert-scale items (Boring and Philippe, 2021), and the SPLE is designed to benefit from this effect. The evidence is less encouraging for open-ended responses. Owen, De Bruin, and Wu (2024) found that structured prompts improved the specificity and constructiveness of open-ended comments but did not reduce gender bias — women faculty were penalized at similar rates across all conditions. This finding is consistent with the broader literature documenting that the unstructured format of open-ended responses gives bias channels that structured items constrain.

Chapter 4

Scoring and Reporting Guidelines

The recommendations in this chapter were approved by the committee with 7 votes in favor, 0 votes against, and 2 abstentions.

4.1 Scoring Methodology

The following scoring approach applies to all items in the **Student Perceptions of Learning Experience.**

- **Ordered categorical data.** The responses are ordered categorical: the categories have a natural ranking but the distances between them are undefined. They are not interval-scale measurements (Stevens, 1946; Jamieson, 2004). The instrument uses a structured fixed-response format — what the Collective Bargaining Agreement terms "Scantron form, etc." (CBA §15.17), and the resulting survey data constitute student course evaluations under that provision.
- **Five ordered categorical response options:** Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree.
- **A Not Applicable (N/A) option** is also available for each question.
- **No numerical scoring.** The categorical responses are not assigned numerical values, as those values cannot be interpreted and their presence encourages misinterpretation. As Stark explains:

  > "While it is common to replace the category names with numbers, for instance, using '1' to signify 'strongly disagree' and '5' to signify 'strongly agree,' the numbers themselves are not quantities, just new labels. They are codes that happen to be numerical. The actual magnitudes of the numbers do not mean anything. The labels are arbitrary. Averaging such numbers is meaningless as a matter of statistics. For the average to be meaningful, the difference between '1' and '2' would need to mean the same thing as

the difference between '4' and '5.' A '1' would have to balance a '5' to be the equivalent of two '3's. But adding or subtracting labels from each other does not make sense, any more than it makes sense to add or average postal codes" (Stark, 2016, ¶28–29; see also Stark, 2026).

4.1.1 Why frequency distributions are preferred over measures of central tendency for this instrument

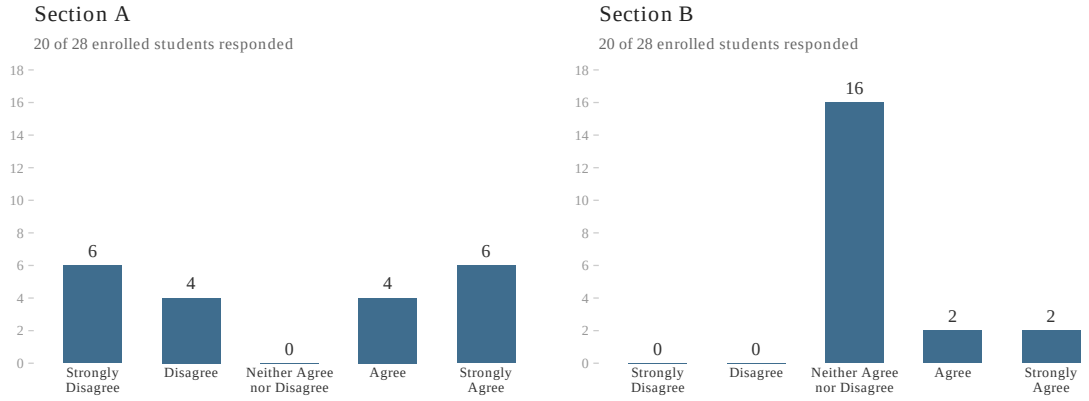
With only five possible values and many more than five students in a classroom, the median cannot move until the distribution shifts enough to push the 50th percentile across a category boundary. Small but meaningful differences — and even some large ones — are invisible to it. This creates two distinct problems.

4.1.1.1 Problem 1: The median hides variation

Two distributions can have very different spreads yet produce the same median.

Table 4.1: Frequency distribution of responses (20 of 28 enrolled students responded)

	Section A	Section B
Strongly Agree	6	2
Agree	4	2
Neither Agree nor Disagree	0	16
Disagree	4	0
Strongly Disagree	6	0



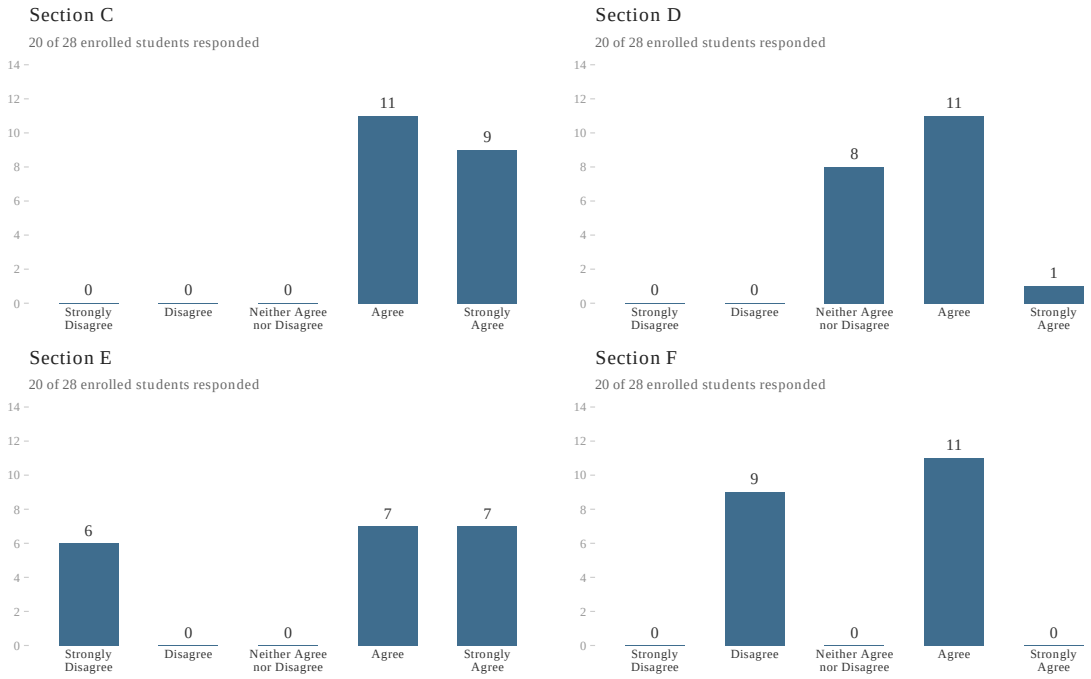
Sections A and B above have the same median (Neither Agree nor Disagree), and very different distributions. Section A is deeply polarized — students are split between strong agreement and strong disagreement. Section B is concentrated at the center. An evaluator seeing "Neither Agree nor Disagree" twice would assume these are similar. They are not.

4.1.1.2 Problem 2: The median is too coarse to locate the center

The median also fails to distinguish distributions that differ in where their weight sits. Problem 1 showed that two distributions with different spreads can share a median. The following examples show that even distributions with very different centers of gravity — where one class is overwhelmingly positive and another is split down the middle — can produce the same median.

Table 4.2: Frequency distribution of responses (20 of 28 enrolled students responded)

	Sec. C	Sec. D	Sec. E	Sec. F
Strongly Agree	9	1	7	0
Agree	11	11	7	11
Neither Agree nor Disagree	0	8	0	0
Disagree	0	0	0	9
Strongly Disagree	0	0	6	0



All four sections report median = "Agree." But Section C is overwhelmingly positive, Section D is lukewarm, Section E is fairly polarized, and Section F is a knife-edge split. The median cannot tell them apart because five categories do not give it enough resolution — the distribution must shift *a lot* before the median moves to the next step. The distributions shown above shift *plenty* and the median does not budge.

The frequency distribution tells you instantly which case you are looking at. The median

hides it.

In practice, the problem is sharper still. Student evaluations are typically such that most students who respond to the survey report nominally positive experiences. With a five-category scale and typical class sizes (15–40 students), the median will almost always fall at "Agree" or "Strongly Agree." This compresses nearly all instructors into two bins, making the median nearly useless for the purpose it is most needed for: helping evaluators distinguish between cases.

A well-designed bar chart is not merely an illustration — it is itself the most effective summary available. As Tufte (1983) observed, the best statistical graphics communicate the full distribution of the data at a glance, rendering patterns, extreme values, and variation instantly legible in a way that no single summary statistic can. For a five-category ordinal variable, a bar chart *is* the summary measure — one that preserves the distributional information that the median discards.

4.2 Reporting Guidelines

- **Frequency distributions.** The number of students whose response falls in each category should be reported as raw counts together with percentages.

> ⚠ Care in interpretation of percentages
>
> Percentages make it easier to compare the shape of a distribution across sections with different numbers of respondents — "30% Strongly Agree" is immediately interpretable in a way that "7 out of 23" requires mental arithmetic. For evaluators reviewing many courses, percentages provide a quicker read of the distributional pattern.
> However, with the class sizes typical of most courses, percentages create a misleading impression of precision: a single student's response can shift a percentage by several points, and the small denominator is hidden from the reader. Reporting counts — e.g., "7 out of 23 respondents" — keeps the sample size visible and discourages over-interpretation (Lang and Secic, 2006, Ch. 1). For this reason, percentages should always be reported alongside raw counts and the total number of respondents, never in isolation.

- **Response rates.** Both the number of enrolled students and the number of respondents should be reported.
- **No extrapolation.** Results should not be extrapolated from responders to nonresponders. Students who submit evaluations are a self-selected sample of convenience, not a random sample; standard statistical measures of uncertainty such as standard errors and confidence intervals are therefore inapt (Stark, 2026).
- **No cross-comparisons.** Results should not be compared across instructors, courses, departments, or disciplines. This is so for the following two reasons:

First, student experience scores are confounded with variables unrelated to teaching effectiveness — including the instructor's gender, race, and age — and these biases are large enough to cause more effective instructors to receive lower scores than less effective ones (Boring, Ottoboni, and Stark, 2016). The bias cannot be corrected because it varies by discipline, by student gender, by survey item, and by other factors. This means that comparing Instructor A's scores to Instructor B's scores — even for the same course — does not reveal who taught more effectively. It reveals the combined effect of demographics, student biases, and nonresponse patterns.

Second, cross-comparisons are invalidated by differences in course characteristics that have nothing to do with teaching: class size, course level, whether the course is required or elective, and student preparation (Stark and Freishtat, 2014, Recommendation 5; McKeachie, 1997, p. 1222). Evaluators should assess each faculty member individually; evaluations and decisions should stand alone without reference to other faculty members or to a unit average (Linse, 2017).

The following table illustrates the recommended reporting format. Each cell contains the raw count and percentage of respondents selecting that category. The table caption states both the number of respondents and the number of enrolled students, making the response rate immediately visible. No numerical averages or information about other instructors or groups of instructors appear.

Table 4.3: Frequency distribution of responses (22 of 33 enrolled students responded; response rate 67 percent)

	Question 1	Question 2
Strongly Agree	6 (27%)	6 (27%)
Agree	7 (32%)	8 (36%)
Neither Agree nor Disagree	4 (18%)	3 (14%)
Disagree	3 (14%)	4 (18%)
Strongly Disagree	2 (9%)	1 (5%)

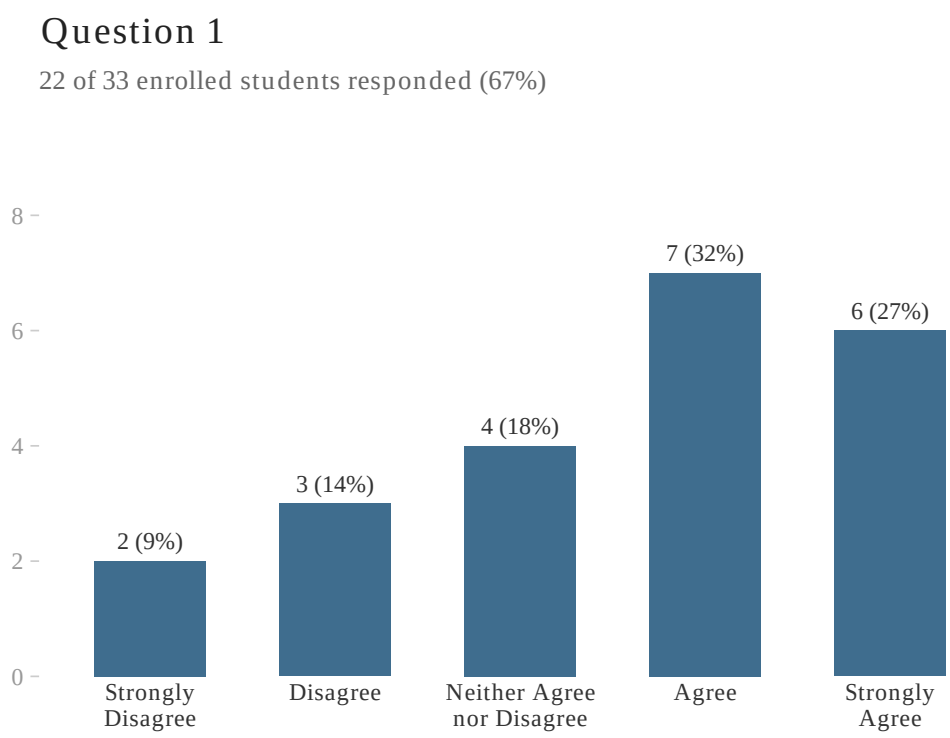
4.3 Visualization Guidelines

The distribution of responses should be examined across the entire scale, not reduced to a single summary statistic (Linse, 2017; Stark and Freishtat, 2014). The distribution should also be displayed as a bar chart showing the count and percentage of respondents in each category.

4.3.1 Bar chart for a single question

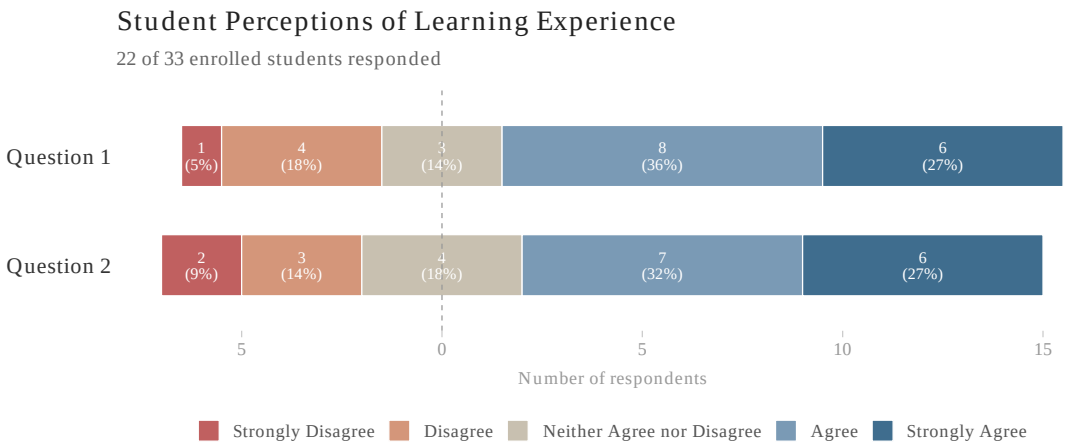
For individual instructor reports, a simple vertical bar chart is the most transparent format. Each bar represents one response category; the vertical axis shows the count of respondents,

with percentages displayed alongside. The response rate appears as a subtitle.



4.3.2 Diverging stacked bar chart for comparing multiple questions

When multiple survey items from an individual need to be compared at a glance, a diverging stacked bar chart is recommended. In this design, proposed by Heiberger and Robbins as "the primary graphical display technique for Likert scales," bars diverge from the neutral midpoint: agreement categories extend to the right, disagreement categories extend to the left, and the neutral category is split evenly across both sides (Heiberger and Robbins, 2014). This layout makes the balance between agreement and disagreement immediately visible — the reader can judge the overall sentiment by comparing the visual mass on each side of the center line. Each segment is labeled with the raw count and percentage; zero-count categories are omitted.


Student Perceptions of Learning Experience
22 of 33 enrolled students responded
Question 1
1 (5%)
4 (18%)
3 (14%)
8 (36%)
6 (27%)
Question 2
2 (9%)
3 (14%)
4 (18%)
7 (32%)
6 (27%)
5
0
5
10
15
Number of respondents
Strongly Disagree
Disagree
Neither Agree nor Disagree
Agree
Strongly Agree

Chapter 5

Implementation Best Practices

The recommendations in this chapter were approved unanimously by the committee.

The preceding chapters define *what* the **Student Perceptions of Learning Experience** (SPLE) measures — six aspects of the learning environment that students are qualified to report on — and *how* its results should be scored and reported. This companion chapter addresses a third question: *how should the instrument be administered?*

Cal Poly's transition to semesters — from 10-week quarters — is a once-in-a-generation opportunity to design the administration of this instrument from scratch rather than inheriting the practices of a system it replaces. The recommendations below draw on the peer-reviewed literature and on the published practices of peer institutions to propose a concrete implementation model for the SPLE.

5.1 Scope

This chapter addresses the implementation of the SPLE — the summative instrument whose results enter the personnel file under CBA §15.17. It does not address the evaluation of teaching more broadly, nor does it address course design, pedagogy, or the other dimensions of the TEval framework that are assessed through peer review, self-reflection, and other evidence sources.

The broader literature on teaching evaluation recognizes that many institutions complement their end-of-term summative instrument with informal mid-semester formative feedback — brief, anonymous check-ins designed to give instructors actionable information while the course is still in progress. Oregon's two-survey model, Angelo and Cross's Classroom Assessment Techniques (1993), and Harvard's early-feedback recommendations (Bok Center) all exemplify this practice. Developing a formative feedback process at Cal Poly is a **separate effort**. A sub-committee of this Ad Hoc Committee prepared a separate document that is not part of this report — Formative Learning Feedback: A Companion to the Student Perceptions of Learning Experience Report — that addresses this topic in detail. This chapter

does not address it further.

The sections that follow focus exclusively on the SPLE instrument: when to administer it, how to administer it, how to maximize response rates, and how to frame it to minimize bias.

5.2 Timing

5.2.1 The literature consensus

The peer-reviewed literature is clear on one point: summative course evaluations should be administered during the **last one to two weeks of instruction, before final examinations begin**. Administering evaluations before students receive final grades avoids contaminating responses with grade-related anxiety or gratitude — a documented source of bias (Marsh, 2007). Administering them too early misses late-semester developments in the learning environment.

What peer institutions do

Institution	Evaluation Window	Source
San José State	~10 days; last 2 weeks of classes	SJSU Teaching Evaluation FAQs for Students; SOTE Interpretation Guide (2022)
San Diego State	~14 days; two-week window before finals	SDSU Student Feedback
UC Davis	Last week of each quarter (~7 days)	UC Davis ACE
UC Santa Barbara	Week 9 Monday – Week 10 Friday (~10 days)	UCSB Course Evaluations
UC San Diego	Week 9 Monday – Week 10 Saturday 8 AM (~6 days)	UCSD SET Faculty FAQ

All of these institutions release results only after final grades have been submitted — a universally recommended practice that protects anonymity and ensures that neither students nor instructors face grade-related pressure during the evaluation period.

5.2.2 Recommendation for Cal Poly's semester

The SPLE window should be **open during the last two weeks of instruction before finals week**. This two-week window is consistent with the practice at most peer institutions, provides sufficient time for reminders and in-class completion, and ensures that the evaluation captures students' experience of nearly the full semester without bleeding into the final

examination period.

5.3 Mode of administration

5.3.1 The response-rate problem

The single most important administrative decision is the *mode* of administration, because it largely determines the response rate.

The evidence is unambiguous: **in-class administration** produces the highest response rates. Paper-based in-class administration historically achieved **80–90%** response rates (Nulty, 2008; Berk, 2012). By contrast, **online-only outside-class** administration typically produces **30–60%** response rates — a range in which self-selection bias is a serious threat to the validity of the data (see Section 5.4).

5.3.2 The hybrid model

A growing number of institutions have adopted a hybrid approach: dedicating class time for students to complete the evaluation *online*, on their own devices. This combines the response-rate benefits of in-class administration with the logistical efficiency of an online platform. Studies report response rates of **70–80%** with this model — comparable to traditional paper-based in-class administration (Berk, 2012; Chapman and Joines, 2017).

The hybrid model is particularly well suited to the SPLE. The instrument is designed to be short and focused — a student can complete it in under ten minutes on a phone. Ten to fifteen minutes of dedicated class time is more than sufficient, even accounting for the time to display the link, wait for students to access it, and allow for thoughtful responses.

5.3.3 Recommendation

The summative SPLE should use a **hybrid model**: during the evaluation window, each instructor dedicates **10–15 minutes of class time** for students to complete the survey online. The instructor displays the survey link (URL or QR code), then **leaves the room**. A designated student or TA signals the instructor to return when time is up. Students who are absent during the in-class session complete the evaluation outside of class during the remainder of the window.

This is the single most effective step the university can take to ensure that the SPLE produces high enough response rates.

5.4 Maximizing response rates

5.4.1 Why response rates matter

When response rates are low, the students who choose to respond may differ systematically from those who do not — they may be more satisfied, more dissatisfied, higher-performing,

or lower-performing than the class as a whole. This self-selection bias is not a theoretical concern; it is well documented. As Stark (2026) emphasizes, students who submit evaluations are a self-selected sample of convenience, not a random sample, and there is no statistical basis for extrapolating from respondents to the class as a whole.

Holtgraves and colleagues (2023) found that non-respondents were not a random subset of enrolled students, that respondents differed from non-respondents in systematic ways, and that the resulting bias could not be corrected by statistical adjustment.

i Strategies ranked by effectiveness: what the literature says

The literature identifies the following strategies, roughly ranked by their demonstrated impact on response rates:

1. **Dedicate class time for online completion.** This is the single most effective intervention. It converts the evaluation from a task students must remember to do on their own time into one that is built into the structure of the course (Berk, 2012; Chapman and Joines, 2017).
2. **Multiple automated reminders.** Adams and Umbach (2012) found that four reminders spaced at 2–3 day intervals brought response rates to approximately the 70th percentile of course-level rates. Each additional reminder (up to four) produced a statistically significant increase.
3. **LMS integration.** Embedding the evaluation link within Canvas — as a dashboard notification, a pop-up reminder, or a course navigation item — reduces the friction of locating and accessing the survey. Students are already in the LMS daily; the evaluation should meet them there.
4. **Instructor communication.** When instructors discuss the evaluation on Day 1 (e.g., a syllabus note explaining that the SPLE asks about the student's learning experience and that the data are read and taken seriously), and again when the evaluation window opens, response rates increase modestly. The mechanism is legitimacy: students participate when they believe their feedback matters (Chen and Hoshower, 2003).
5. **Class-level incentives.** Goodman, Anson, and Belcheir (2015) found that a class-level incentive (e.g., a bonus point if the class achieves an 80% response rate) increased response rates by approximately 22 percentage points. Class-level incentives avoid the coercion problem of individual incentives because no individual student's participation can be identified.

5.4.2 Recommendation

We recommend the following:

- **Dedicate class time.** Each instructor should set aside 10–15 minutes during the evaluation window for students to complete the SPLE online in class. The instructor

displays the survey link or QR code, then leaves the room. This is the single most effective intervention for achieving high response rates.

- **Send four automated reminders** at 2–3 day intervals during the evaluation window, via email and Canvas notification.
- **Integrate with the LMS.** Embed the evaluation link within Canvas — as a dashboard notification, pop-up reminder, or course navigation item — so that the survey meets students where they already are.
- **Encourage instructor communication.** A brief mention on Day 1 (e.g., a syllabus note explaining that the SPLE asks about the student's learning experience and that the data are taken seriously), repeated when the window opens, increases participation.

5.5 Framing the instrument to minimize bias

5.5.1 The evidence on anti-bias framing

A natural question is whether the instructions presented to students before they complete the evaluation can reduce the biases documented in the literature — particularly gender bias. The answer is nuanced: it depends entirely on what kind of framing is used.

Normative framing — generic appeals to fairness such as "Please evaluate your instructor fairly, regardless of their gender, race, or other characteristics" — has been shown to have **no significant effect** on evaluation outcomes. Boring and Philippe (2021) tested this directly in a large-scale field experiment at Sciences Po and found that a normative anti-bias warning produced no detectable change in the gender gap.

Informational framing — pairing the warning with institution-specific data showing that previous cohorts had evaluated male and female instructors differently — produced a markedly different result. In the same experiment, Boring and Philippe found that informational framing significantly reduced the gender bias, raising ratings of female instructors without affecting ratings of male instructors. The effect was driven primarily by male students' evaluations of female instructors; female students' ratings were not significantly affected by either treatment.

An important caveat: The evidence that informational framing reduces bias applies to *structured Likert-scale items*. It does not extend to *open-ended responses*, where the unstructured format gives bias room to operate regardless of how the prompt is framed. Owen, De Bruin, and Wu (2024) found that even directed, structured prompts — while they improved the specificity and constructiveness of open-ended comments — did not reduce gender bias. This is one of the reasons the committee considered discontinuing open-ended questions from the summative instrument, and ultimately voted to retain them only under the structured prompts and guardrails described in Chapter 3.

5.5.2 Recommendation

The SPLE should open with a brief, concrete, **data-informed preamble** — not a generic "be fair" appeal, which the evidence shows is ineffective, but a factual statement that provides students with context about what the survey measures and what the research shows about evaluation biases. The preamble should:

1. **Name what the survey measures.** Remind students that the SPLE asks about their *own experience* of the learning environment — not a verdict on the instructor as a person or professional.
2. **Provide specific information about documented biases.** A brief, factual statement — e.g., "Research shows that students' evaluations of their learning experience can be influenced by characteristics of the instructor unrelated to the learning environment, such as gender and race. Being aware of this tendency helps produce more accurate feedback."
3. **Reinforce the survey's purpose.** The data are used to understand the student learning experience and to support faculty development and evaluation. Thoughtful, honest responses improve the quality of the data.

The name **Student Perceptions of Learning Experience** is itself a framing device. By directing attention to the student's *experience* rather than to the instructor's *performance*, the instrument's name reinforces the experiential focus that the bias literature recommends.

> **i** Draft preamble language
>
> **Student Perceptions of Learning Experience**
> This brief survey asks about *your experience* in this course — the learning environment, your interactions with the instructor, and how you perceive the course was structured. It does not ask you to evaluate the instructor's teaching ability or the course content.
> Research shows that students' responses to surveys like this can be influenced by characteristics of the instructor — such as gender, race, and accent — that are unrelated to the learning environment. Being aware of this tendency helps you provide more accurate feedback.
> Your responses are anonymous and will not be shared with the instructor until after final grades have been submitted. Please respond thoughtfully and honestly.

5.6 Recommended implementation model

The following table synthesizes the evidence reviewed in this chapter into a concrete recommendation for administering the summative Student Perceptions of Learning Experience under Cal Poly's semester calendar.

Element	Recommendation	Rationale
When	Last two weeks of instruction before finals week	Literature consensus: last 1–2 weeks of instruction, before finals begin (Marsh, 2007)
Mode	Hybrid: dedicated class time for online completion	Single most effective method for achieving 70%+ response rates (Berk, 2012)
Class time	10–15 min; instructor displays link/QR code, then leaves	The instrument is short and focused — feasible in under ten minutes
Framing	Informational preamble (data-informed, not generic)	Boring and Philippe (2021): informational framing reduces gender bias; generic appeals do not
Reminders	4 automated reminders at 2–3 day intervals	Adams and Umbach (2012): achieves ~70% response rates
Results release	Only after final grades are submitted	Universally recommended; protects anonymity and reduces grade-anxiety bias

Administering the SPLE in the final two weeks of instruction, with dedicated class time and an informational preamble is an achievable model — it requires no new technology, no additional personnel, and minimal class time — and it reflects the best available evidence on how to implement a student survey that is both useful and fair.

5.7 Open implementation questions

The implementation considerations discussed in this chapter are not exhaustive. During the Academic Senate discussion of this report on May 26, 2026, a faculty member raised concerns about the reliability of the process, including who is eligible to respond, how to handle students who have dropped a course, and how to handle students facing disciplinary action. These are important questions and should be addressed carefully.

There are likely other edge cases and implementation matters that this chapter has not fully considered, including small-enrollment courses, cross-listed courses, and differences across modalities such as in-person, hybrid, and online instruction. The group charged with revising the instrument and developing the implementation plan should take up these issues. Our purpose here is to flag them so that those who inherit this work know to examine them.

Chapter 6

Conclusions

The **Student Perceptions of Learning Experience** proposed here is a short, focused instrument grounded in the peer-reviewed literature and aligned with the UFPP's requirements. By renaming the instrument, centering it on the six aspects of class climate that students are qualified to report on, and adopting guardrails for the open-ended questions to mitigate documented equity bias, we can give students a meaningful voice in the evaluation of teaching while protecting both students and instructors from the well-documented biases of traditional teaching evaluations. This is an achievable reform — one that strengthens the integrity of the evaluation process and brings Cal Poly's practices in line with a growing movement across higher education toward multidimensional, evidence-based evaluation of teaching (McCreary, 2026; Stark, 2026). Moreover, the instrument reflects Cal Poly's distinctive pedagogical identity: at a university where students learn by doing, a survey centered on the climate in which that doing takes place is not just methodologically sound — it is institutionally apt.

In addition, two sub-committees of this Ad Hoc Committee prepared companion documents that are separate from this report. These documents have not been formally adopted by the full committee and are offered as companion resources for consideration by the respective Academic Senate committees to which they are addressed.

- **Guidance for Evaluation of Instruction** — a proposed revision to UFPP §8.3 that organizes the evaluation of teaching around the seven TEval dimensions, includes a teaching effectiveness rubric adapted from the University of Kansas Benchmarks for Teaching Excellence, provides guidance on the appropriate and inappropriate uses of the survey data (including the inherent limitations of student evaluation data and the role of SPLE results within the broader evaluation framework), addresses department-associated questions, sets departmental expectations, and outlines training and implementation requirements. Offered for consideration by the Academic Senate Faculty Affairs Committee.
- **Formative Learning Feedback: A Companion to the Student Perceptions of Learning**

Experience Report — a voluntary, developmental feedback process to be tentatively offered through the Center for Teaching, Learning and Technology (CTLT), organized around the seven research-based principles of learning identified by Ambrose et al. (2010), and designed to give instructors actionable information about how students are experiencing the learning environment while the course is still in progress. Offered for consideration by the Academic Senate Instruction Committee and the Center for Teaching, Learning and Technology.

References

Abrami, P. C. (2001). Improving judgments about teaching effectiveness using teacher rating forms. *New Directions for Institutional Research*, 2001(109), 59–87. https://doi.org/10.1002/ir.4

Adams, M. J. D., & Umbach, P. D. (2012). Nonresponse and online student evaluations of teaching. *Research in Higher Education*, 53(2), 153–168. https://doi.org/10.1007/s11162-011-9240-5

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How Learning Works: Seven Research-Based Principles for Smart Teaching*. Jossey-Bass.

Austin, A. E., Finkelstein, N. D., Greenhoot, A. F., Ward, D., & Weaver, G. C. (2025). *Transforming College Teaching Evaluation: A Framework for Advancing Instructional Excellence*. Harvard Education Press. https://hep.gse.harvard.edu/9798895570159/transforming-college-teaching-evaluation/

Benton, S. L., & Young, S. (2018). Best practices in the evaluation of teaching. *IDEA Paper No. 69*. https://eric.ed.gov/?id=ED588352

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International Journal of Teaching and Learning in Higher Education*, 17(1), 48–62. https://www.isetl.org/ijtlhe/pdf/IJTLHE8.pdf

Berk, R. A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching and Learning*, 8(2), 98–107. https://apps.weber.edu/wsuimages/ie/Evaluations/Top%2020%20strategies%20to%20increase%20online%20response%20rates.pdf

Boring, A., Ottoboni, K., & Stark, P. B. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*. https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Boring, A., & Philippe, A. (2021). Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching. *Journal of Public Economics*, 193, 104323. https://doi.org/10.1016/j.jpubeco.2020.104323

Boysen, G. A., Kelly, T. J., Raesly, H. N., & Casner, R. W. (2014). The (mis)interpretation of teaching evaluations by college faculty and administrators. *Assessment & Evaluation in*

Higher Education, 39(6), 641–656. https://eric.ed.gov/?id=EJ1030222

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

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

Centra, J. A. (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness*. Jossey-Bass. https://eric.ed.gov/?id=ED363233

Chapman, D. D., & Joines, J. A. (2017). Strategies for increasing response rates for online end-of-course evaluations. *International Journal of Teaching and Learning in Higher Education*, 29(1), 47–60. https://doi.org/10.1007/s10755-017-9394-0

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment & Evaluation in Higher Education*, 28(1), 71–88. https://doi.org/10.1080/02602930301683

Chisadza, C., Nicholls, N., & Yitbarek, E. (2019). Race and gender biases in student evaluations of teachers. *Economics Letters*, 179, 66–71. https://doi.org/10.1016/j.econlet.2019.03.022

Cho, W., Baek, W., & Cho, J. (2015). Why do good performing students highly rate their instructors? Evidence from a natural experiment. *Economics of Education Review*, 49, 172–179. https://doi.org/10.1016/j.econedurev.2015.10.001

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, 116(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Feeley, T. H. (2002). Evidence of halo effects in student evaluations of communication instruction. *Communication Education*, 51(3), 225–236. https://doi.org/10.1080/03634520216519

Fraser, B. J. (1998). Classroom environment instruments: Development, validity and applications. *Learning Environments Research*, 1, 7–34. https://doi.org/10.1023/A:1009932514731

Fraser, B. J., Treagust, D. F., & Dennis, N. C. (1986). Development of an instrument for assessing classroom psychosocial environment at universities and colleges. *Studies in Higher Education*, 11(1), 43–54. https://doi.org/10.1080/03075078612331378451

Frisby, B. N., & Martin, M. M. (2010). Instructor–student and student–student rapport in the classroom. *Communication Education*, 59(2), 146–164. https://doi.org/10.1080/

03634520903564362

Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2022). Reflective writing analytics for actionable feedback. *Australasian Journal of Educational Technology*, 38(1). https://ajet.org.au/index.php/AJET/article/view/6133

Goodman, J., Anson, R., & Belcheir, M. (2015). The effect of incentives and other instructor-driven strategies to increase online student evaluation response rates. *Assessment & Evaluation in Higher Education*, 40(7), 958–970. https://doi.org/10.1080/02602938.2014.960364

Hagenauer, G., & Volet, S. E. (2014). Teacher–student relationship at university: An important yet under-researched field. *Oxford Review of Education*, 40(3), 370–388. https://doi.org/10.1080/03054985.2014.921613

Hartwig, M. K., & Dunlosky, J. (2017). Category learning judgments in the classroom: Can students judge how well they know course topics? *Contemporary Educational Psychology*, 49, 80–90. https://doi.org/10.1016/j.cedpsych.2016.12.002

Heffernan, T. (2023). Abusive comments in student evaluations of courses and teaching: The attacks women and marginalized academics endure. *Higher Education*, 85, 225–239. https://link.springer.com/article/10.1007/s10734-022-00831-x

Heiberger, R. M., & Robbins, N. B. (2014). Design of diverging stacked bar charts for Likert scales and other applications. *Journal of Statistical Software*, 57(5), 1–32. https://doi.org/10.18637/jss.v057.i05

Hessler, M., et al. (2018). Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education*, 52, 1064–1072. https://doi.org/10.1111/medu.13627

Hurtado, S., Alvarez, C. L., Guillermo-Wann, C., Cuellar, M., & Arellano, L. (2012). A model for diverse learning environments. In J. C. Smart & M. B. Paulsen (Eds.), *Higher Education: Handbook of Theory and Research* (Vol. 27). Springer. https://doi.org/10.1007/978-94-007-2950-6_2

Hutchinson, M., Coutts, R., Massey, D., Nasrawi, D., Fielden, J., Lee, M., & Lakeman, R. (2024). Student evaluation of teaching: reactions of Australian academics to anonymous non-constructive student commentary. *Assessment & Evaluation in Higher Education*, 49(2), 154–164. https://www.tandfonline.com/doi/full/10.1080/02602938.2023.2195598

Knof, H., Berndt, M., & Shiowaza, T. (2024). Prevalence of Dunning-Kruger effect in first semester medical students. *BMC Medical Education*, 24, 1210. https://doi.org/10.1186/s12909-024-06121-7

Kreitzer, R. J., & Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in SETs and recommendations for ethical reform. *Journal of Academic Ethics*, 20, 73–84. https://link.springer.com/article/10.1007/s10805-021-09400-w

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Lake, D. A. (2001). Student performance and perceptions of a lecture-based course compared with the same course utilizing group discussion. *Physical Therapy*, 81(3), 896–902. https://academic.oup.com/ptj/article-abstract/81/3/896/2857593

Lang, T. A., & Secic, M. (2006). *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers* (2nd ed.). American College of Physicians. https://openlibrary.org/works/OL18654342W

Lindsey, B. A., & Nagel, M. L. (2015). Do students know what they know? Exploring the accuracy of students' self-assessments. *Physical Review Special Topics — Physics Education Research*, 11, 020103. https://doi.org/10.1103/PhysRevSTPER.11.020103

Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94–106. https://doi.org/10.1016/j.stueduc.2016.12.004

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes. *Studies in Higher Education*, 27(1), 27–52. https://doi.org/10.1080/03075070120099359

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, 40, 291–303. https://doi.org/10.1007/s10755-014-9313-4

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11(3), 253–388. https://doi.org/10.1016/0883-0355(87)90001-2

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education*. Springer. https://doi.org/10.1007/1-4020-5742-3_9

McCreary, M. (2026). A practical guide to modern teaching evaluation. *Engaged Learning Collective*. https://engagedlearningcollective.substack.com/p/a-practical-guide-to-modern-teaching-evaluation

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218–1225. https://doi.org/10.1037/0003-066X.52.11.1218

Mengel, F., Sauermann, J., & Zölitz, U. (2018). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535–566. https://doi.org/10.1093/jeea/jvx057

Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648–652. https://www.cambridge.org/core/journals/ps-political-

science-and-politics/article/gender-bias-in-student-evaluations/1224BE475C0AE75A2C2D8553210C4E27

Moos, R. H. (1979). *Evaluating Educational Environments*. Jossey-Bass. https://cir.nii.ac.jp/crid/1970304959869740454

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys. *Assessment & Evaluation in Higher Education*, 33(3), 301–314. https://doi.org/10.1080/02602930701293231

Owen, A. L., De Bruin, E., & Wu, S. (2024). Can you mitigate gender bias in student evaluations of teaching? Evaluating alternative methods of soliciting feedback. *Assessment & Evaluation in Higher Education*, 50(3). https://doi.org/10.1080/02602938.2024.2407927

Rivera, L., & Tilcsik, A. (2019). Scaling down inequality: Rating scales, gender bias, and the architecture of evaluation. *American Sociological Review*, 84(2), 248–274. https://doi.org/10.1177/0003122419833601

Stark, P. B. (2016). *An evaluation of course evaluations*. Report for the Task Force on Assessing Teaching. https://www.tfanet.ca/wp-content/uploads/2018/11/Stark_report.pdf

Stark, P. B. (2026). Using SET for employment decisions is unethical. SSRN Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6193538

Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Storage, D., Horne, Z., Cimpian, A., & Leslie, S.-J. (2016). The frequency of "brilliant" and "genius" in teaching evaluations is correlated with the representation of women and African Americans across fields. *PLoS ONE*, 11(3), e0150194. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150194

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation. *Basic and Applied Social Psychology*, 42(4), 276–294. https://doi.org/10.1080/01973533.2020.1756817

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Graphics Press. https://www.edwardtufte.com/book/the-visual-display-of-quantitative-information/

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007

Wagner, N., Rieger, M., & Voorvelt, K. (2016). Gender, ethnicity and teaching evaluations: Evidence from mixed teaching teams. *Economics of Education Review*, 54, 79–94. https://doi.org/10.1016/j.econedurev.2016.06.004

Williams, W. M., & Ceci, S. J. (1997). "How'm I doing?": Problems with student ratings of instructors and courses. *Change: The Magazine of Higher Learning*, 29(5), 12–23. https://doi.org/10.1080/00091389709602331

Woolliscroft, J. O., TenHaken, J., Smith, J., & Calhoun, J. G. (1993). Medical students' clinical self-assessments: comparisons with external measures of performance and the students' self-assessments of overall performance and effort. *Academic Medicine*, 68(4), 285–294. https://pubmed.ncbi.nlm.nih.gov/8466612/

Xu, X., et al. (2024). Differences between resident self-assessments and faculty-assessments on Anesthesiology Milestones and associated factors. *BMC Medical Education*, 24, 551. https://doi.org/10.1186/s12909-024-05544-6

Appendix A

Appendix: Sample Survey Instrument

This appendix presents a sample version of the **Student Perceptions of Learning Experience** (**SPLE**), including the recommended preamble and one item per aspect of class climate (two for Coherence). The instrument uses a five-point ordered categorical scale (Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, Strongly Disagree) plus a Not Applicable option for each item. The items presented here are illustrative. They are intended to demonstrate how the six aspects of class climate can be operationalized as experiential survey items. **This is not intended to be the final instrument.**

A.1 Preamble

The survey should open with the following informational preamble, consistent with the evidence on anti-bias framing discussed in Section 5.5 (Boring and Philippe, 2021).

> **Student Perceptions of Learning Experience**
> This brief survey asks about *your experience* in this course — the learning environment, your interactions with the instructor, and how you perceive the course was structured. It does not ask you to evaluate the instructor's teaching ability or the course content.
> Research shows that students' responses to surveys like this can be influenced by characteristics of the instructor — such as gender, race, and accent — that are unrelated to the learning environment. Being aware of this tendency helps you provide more accurate feedback.
> Your responses are anonymous and will not be shared with the instructor until after final grades have been submitted. Please respond thoughtfully and honestly.

A.2 Sample Items

All items use the following response scale:

Strongly Agree · Agree · Neither Agree nor Disagree · Disagree · Strongly Disagree · Not Applicable

A.2.1 Regard for Students

"I felt the instructor engaged with students as individuals."

A.2.2 Consistent Communication and Enforcement of Expectations

"I knew what was expected of me in this course."

"I felt the instructor applied the same expectations and standards to all students."

A.2.3 Access to Instructor and Instructor Resources

"I was able to get help from my instructor when I needed it (in office hours, after class, or by email)."

"I was able to access the course materials and resources I needed for this class."

A.2.4 Perceived Course Coherence

"I could see how what was assessed related to what was covered in the course."

"I could see how the different parts of this course fit together."

A.2.5 Participatory Climate

"I felt there were ways for me to participate in the course."

"I felt the instructor created opportunities for me to explore the ideas in the course."

A.2.6 Responsive Learning Environment

"I felt the instructor created a learning environment that was responsive to all students."

A.3 Sample if the Academic Senate elects to retain open-ended questions

If open-ended questions are retained under the guardrails described in Section 3.3, the instrument would include one structured open-ended prompt on Perceived Course Coherence — the aspect where elaboration is most informative and least susceptible to bias. The prompt appears immediately after the Perceived Course Coherence Likert items and directs the student to describe their experience with course structure.

A.3.1 Perceived Course Coherence (with structured open-ended prompt)

"I could see how what was assessed related to what was covered in the course."

"I could see how the different parts of this course fit together."

"Please describe your experience with how the different parts of this course fit together — for example, how readings, class activities, assignments, and assessments related to each other. Focus on specific aspects of the course, not on personal characteristics of the instructor."

Your response here...

All other items (Regard for Students, Consistent Communication and Enforcement of Expectations, Access to Instructor and Instructor Resources, Participatory Climate, Responsive Learning Environment) remain Likert-only.

A.3.2 Why an open-ended question only on Perceived Course Coherence?

The committee considered attaching an open-ended prompt to each of the six aspects and concluded that Perceived Course Coherence is the only aspect where the benefit of elaboration clearly outweighs the risk of bias. The reasoning, aspect by aspect:

- **Regard for Students.** An open-ended prompt here invites commentary on manner, demeanor, and personality — exactly the content that disproportionately targets women and faculty from marginalized groups (Mitchell and Martin, 2018). Highest risk, lowest benefit.
- **Consistent Communication and Enforcement of Expectations.** An open-ended prompt here invites comments about grading, which correlates with grade *expectations*, not actual consistency of standards. It also invites favoritism allegations that can be racialized — and open-ended comments are precisely where equity bias is most evident (Kreitzer and Sweet-Cushman, 2021). High risk.
- **Access.** An open-ended prompt here invites commentary on communication style, accent, and warmth — all heavily gendered and racialized (Subtirelu, 2015; Miller and Chamberlin, 2000). High risk.
- **Responsive Learning Environment.** An open-ended prompt here could elicit valuable information, but it could also produce comments about the instructor's identity that are impossible to disentangle from bias. A student who doesn't feel they belong might attribute it to the instructor's demographics rather than to specific practices (Heffernan, 2023). Moderate-to-high risk.
- **Perceived Course Coherence.** This is the safest choice. An open-ended prompt here channels comments toward course structure — readings, assignments, assessments,

the connections between topics. These are the most impersonal, practice-oriented comments a student can make. It is hard (but not impossible) to write something biased about whether the exam matched the lectures. And it is the aspect where elaboration is most useful to evaluators — a Likert response tells you the student didn't see the connections; a structured comment tells you *which* connections were missing.

- **Participatory Climate.** An open-ended prompt here could produce useful structural feedback (e.g., "group work was dominated by two people," "questions were welcomed but never answered"). But it readily invites evaluative commentary about the instructor's teaching style — particularly judgments like "the lectures were boring" or "there was too much group work." Research shows that students conflate instructor enthusiasm and charisma with teaching effectiveness, even though enthusiasm is not associated with learning (Feeley, 2002; Williams and Ceci, 1997). A comment like "boring" tells you about the student's affective response — which may reflect the instructor's gender, accent, or presentation style — not about whether the environment supported participation. This kind of feedback is valuable in the formative process, where the instructor can contextualize it; in the personnel file, it becomes indistinguishable from bias. Moderate risk.

A.4 Relationship to Existing Cal Poly Maritime Academy Practices

Several of these aspects are already tracked in other CSU instruments. The Cal Poly Maritime Academy, for example, includes items on Consistent Communication and Enforcement of Expectations ("The instructor attempted to be fair and unbiased in their interaction with students"), Responsive Learning Environment ("The instructor demonstrated awareness and consideration of the diversity of students in the class"), Access to Instructor and Instructor Resources ("The instructor was responsive when I had questions"), and Participatory Climate ("The instructor provided opportunities for class participation"). The SPLE items are compatible with this existing practice. The principal difference is one of framing: the SPLE items are worded as first-person experiential reports ("I felt…") rather than third-person assessments of instructor behavior ("The instructor attempted…"), consistent with the evidence that experiential items are less susceptible to bias than evaluative ones.

Formative Learning Feedback

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

2026-05-07

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 for consideration by the Academic Senate Instruction Committee and the Center for Teaching, Learning and Technology.

This document describes and proposes **Formative Learning Feedback** — a flexible, voluntary feedback process to be potentially offered through the Center for Teaching, Learning and Technology (CTLT) and designed to give instructors actionable information about how students are experiencing the learning environment while the course is still in progress.

Three foundational principles govern the formative feedback process:

a. **Opt-in.** Formative Learning Feedback is entirely voluntary. No instructor is required to use it, and no administrator may mandate its use.

b. **Results only to the instructor.** Feedback is shared only with the instructor. It is not included in the personnel file, not reported to department chairs or deans, and not used for retention, tenure, promotion, or any other employment decision.

c. **Developmental, not evaluative.** The purpose of formative feedback is to help the instructor improve the learning environment in real time. It is not a tool for summative evaluation and should never be framed as one.

Because formative feedback is developmental and shared only with the instructor, it can also be the appropriate home for general-purpose **open-ended questions** — the kind of unstructured, in-their-own-words feedback that is most useful when the instructor can still act on it, and that the literature identifies as too susceptible to bias for inclusion in a summative personnel file.

Formative Learning Feedback is entirely separate from the **Student Perceptions of Learning Experience (SPLE)**, which is the summative instrument whose results enter the personnel file. The two serve fundamentally different purposes and operate under different rules.

Chapter 1

Why Formative Feedback?

1.1 Purpose and scope

This document describes a feedback process designed to give instructors actionable information about the learning environment while the course is still in progress.

The formative feedback process is **developmental only**:

- Results are shared **only with the instructor**
- Results are **not** included in the personnel file
- Results are **not** used for retention, tenure, promotion, or any other employment decision
- Results are **not** reported to department chairs, deans, or any administrative unit

1.2 The case for feedback throughout the term

Traditional mid-semester evaluations offer a single snapshot — typically around weeks 6–8 — that arrives too early for some concerns and too late for others. Students entering a new subject may not have enough experience at week 6 to report on how practice and feedback are supporting their learning, while by week 8 it may be too late to address confusion about how course elements connect.

A more effective model provides **multiple touchpoints** across the term, each targeting the dimensions of learning most relevant at that stage:

- **Early in the term (weeks 2–3):** Students are orienting to the course. Feedback on whether the course is building on what they already know and whether they can see how the pieces fit together is most actionable here.
- **Mid-term (weeks 6–8):** Students have enough experience to report on motivation, the quality of practice and feedback, and the classroom climate.
- **Late in the term (weeks 11–12):** Students can reflect on whether the course has helped

them develop mastery and self-directed learning skills — and there is still time for the instructor to adjust the final weeks.

Faculty choose which touchpoints to use and which dimensions to ask about. The instrument is a menu, not a mandate.

1.3 Evidence base

The formative feedback process draws on a well-established literature on feedback practices in higher education.

1.3.1 Classroom Assessment Techniques (CATs)

Angelo and Cross (1993) developed a comprehensive set of Classroom Assessment Techniques — brief, usually anonymous, in-class activities designed to give instructors rapid feedback on student learning and experience. The most widely used CATs include the Minute Paper, the Muddiest Point, and the One-Sentence Summary. The design principles underlying CATs — brevity, anonymity, low stakes, instructor-initiated — inform the formative feedback process proposed here.

1.3.2 Small Group Instructional Diagnosis (SGID)

The SGID method, developed at the University of Washington (Clark and Redmond, 1982), uses a trained facilitator to gather structured feedback from small groups of students during class time, with the instructor absent. The facilitator synthesizes the responses and meets privately with the instructor. SGID is more resource-intensive than a written check-in but produces richer, more contextualized feedback. Institutions with active SGID programs include Indiana University (Center for Innovative Teaching and Learning), UCLA, and the University of Massachusetts Amherst.

1.3.3 Oregon Mid-Semester Experience Survey (M-SES)

The University of Oregon operates a two-survey model: a formative **Midway Student Experience Survey (M-SES)** and a summative End-of-term Student Experience Survey (E-SES). The M-SES is administered during week 4 of the 10-week quarter (equivalent to approximately weeks 6–7 of a 15-week semester) and asks students about their learning experience to date. Instructors are encouraged to provide 10 minutes of class time for completion. Results are shared only with the instructor (University of Oregon TEP). Notably, Oregon's Senate motion US18/19-14 phased out traditional "course evaluations" in favor of learning-focused "Student Experience Surveys" — and pilot data showed that personal comments about instructors dropped from 21% to 1.5% of all comments under the new instrument. The Oregon model demonstrates that mid-semester feedback can be institutionally supported without being tied to personnel decisions.

1.3.4 Harvard Bok Center

Harvard's Derek Bok Center for Teaching and Learning recommends that instructors gather "early, frequent, and low-stakes feedback" from students, using brief anonymous surveys or structured class discussions. The emphasis is on actionable feedback that can inform adjustments while the course is still in progress (Bok Center).

1.3.5 Indiana University CITL

The Center for Innovative Teaching and Learning at Indiana University offers both SGID facilitation and a mid-semester feedback template that instructors can administer independently. The CITL model emphasizes "closing the loop" — the instructor's public response to the feedback received — as essential to the process's effectiveness (Indiana CITL).

1.3.6 UCLA Center for the Advancement of Teaching

UCLA's Center for the Advancement of Teaching provides mid-semester evaluation resources including facilitated SGID sessions and self-administered survey templates. The program is framed explicitly as a developmental tool with no connection to personnel review (UCLA Center for the Advancement of Teaching).

1.4 Theoretical backbone: Ambrose et al. (2010)

The formative feedback instrument is organized around the seven research-based principles of learning identified by Ambrose et al. (2010) in *How Learning Works: Seven Research-Based Principles for Smart Teaching*. These principles synthesize decades of cognitive and educational research into a practical framework for understanding how students learn. Each principle is framed as a question — *How does students' prior knowledge affect their learning?* — and each suggests specific dimensions of the learning experience that students are positioned to observe and report on.

The mapping of these seven principles to candidate feedback items draws on work by Committee Member Patrick O'Sullivan (CTLT), who identified the teaching practices associated with each principle that students can directly experience and comment on. This mapping also connects the formative dimensions to the TEval framework (Austin et al., 2025), providing a coherent link between the formative feedback process and the broader evaluation of teaching.

The seven dimensions — Prior Knowledge, Knowledge Organization, Motivation, Mastery, Practice and Feedback, Student Development and Course Climate, and Self-Directed Learning — are described in detail in Chapter 2.

Chapter 2

Integrating Effective Teaching Practices Into the Formative Learning Feedback

The formative feedback instrument is organized around seven dimensions of learning, each drawn from the research-based principles identified by Ambrose et al. (2010). For each dimension, we provide:

- The **guiding question** from Ambrose et al.
- A brief description of **what it captures**
- A note on **why the dimension is appropriate formatively but not summatively**

All candidate items to be offered to Faculty are phrased as student-experience statements about observable teaching practices. Faculty select which dimensions and items to include based on their course, their goals, and the timing of the feedback. They can also include their own items.

Foundations — what students bring and how the course builds on it:

1. Prior Knowledge
2. Knowledge Organization

Engagement — what drives and sustains the learning process:

3. Motivation
4. Mastery
5. Practice and Feedback

Environment and autonomy — the climate for learning and beyond:

6. Student Development and Course Climate
7. Self-Directed Learning

2.1 Dimension 1: Prior Knowledge

Guiding question: *How does students' prior knowledge affect their learning?*

What it captures: Whether the instructor creates opportunities for students to connect new material to what they already know — including opportunities to surface and correct misconceptions.

> ! Why formative, not summative?
>
> Whether an instructor effectively activates prior knowledge requires understanding the content and the pedagogical choices involved. Students can report on whether opportunities were provided — a valuable formative signal — but cannot judge whether the strategies were appropriate for the subject matter.

2.2 Dimension 2: Knowledge Organization

Guiding question: *How does the way students organize knowledge affect their learning?*

What it captures: Whether the course helps students see how concepts relate to one another — through explicit organizational frameworks, concept maps, or other structuring activities.

> ! Why formative, not summative?
>
> Judging whether a course's organizational structure is effective requires pedagogical expertise. Students can report on whether organizational support was provided, but the quality and appropriateness of that support depend on disciplinary context that students are not positioned to evaluate summatively.

2.3 Dimension 3: Motivation

Guiding question: *What factors motivate students to learn?*

What it captures: Whether the instructor helps students understand the value and purpose of learning activities, supports students' sense of efficacy, and fosters a climate conducive to engagement.

> ! Why formative, not summative?
>
> Motivation is influenced by many factors beyond the instructor's control — including the student's own goals, preparation, and external circumstances. Students can report on whether the instructor took actions to support motivation, which is valuable formative feedback. But summative evaluation of motivational support risks confounding the instructor's practices with factors outside their influence.

2.4 Dimension 4: Mastery

Guiding question: *How do students develop mastery?*

What it captures: Whether the course provides opportunities for students to acquire, practice, and integrate component skills toward increasingly complex performance.

> ! Why formative, not summative?
>
> Whether a course effectively scaffolds skill development toward mastery requires understanding the disciplinary goals and the appropriateness of the progression. Students can report on whether practice and integration opportunities were provided — actionable feedback for the instructor — but cannot judge whether the progression was well-designed for the learning goals of the course.

2.5 Dimension 5: Practice and Feedback

Guiding question: *What kinds of practice and feedback enhance learning?*

What it captures: Whether students receive goal-directed practice with clear criteria for success, and whether feedback is timely and specific enough to guide improvement.

> ! Why formative, not summative?
>
> The quality and timeliness of feedback is something students can directly experience and report on, making it excellent formative data. However, summative evaluation of feedback practices risks conflating the *experience* of feedback (which may feel harsh or generous regardless of quality) with its *effectiveness* (which requires pedagogical judgment). The well-documented disconnect between perceived and actual learning (Deslauriers et al., 2019) applies directly here.

2.6 Dimension 6: Student Development and Course Climate

Guiding question: *Why do student development and course climate matter for student learning?*

What it captures: Whether the classroom climate promotes a sense of belonging and whether norms for interaction support mutual respect — climate as it supports the learning process.

> ! How Dimension 6 relates to the SPLE
>
> This dimension overlaps with the summative SPLE, which also addresses class climate. The overlap is intentional: climate is important enough to warrant both a formative check (visible only to the instructor, actionable mid-course) and a summative record (entered in the personnel file at term's end). The two serve different institutional

purposes even when they touch the same territory.

2.7 Dimension 7: Self-Directed Learning

Guiding question: *How do students become self-directed learners?*

What it captures: Whether the instructor models and supports metacognitive practices — helping students assess what they know, identify what they still need to learn, and develop strategies for continued learning.

> **!** Why formative, not summative?
>
> Self-directed learning is a developmental outcome that unfolds over time and across courses. Students can report on whether metacognitive support was provided in a specific course — helpful formative feedback — but cannot assess whether those practices were effective in building lasting self-regulation skills.

Chapter 3

Relationship to the Summative SPLE

3.1 Comparison

The formative feedback process is **not** a practice run for the summative SPLE, and it should not be framed as one. The two instruments differ in every relevant dimension:

Table 3.1: Comparison of formative and summative instruments

Dimension	Formative Learning Feedback	Summative SPLE
Purpose	Developmental: help the instructor improve the course in real time	Evaluative: provide data for the personnel file
Audience	Instructor only	Instructor, department chair, personnel committee
Timing	Throughout the term (weeks 2–3, 6–8, 11–12)	Last two weeks of instruction before finals week
Content	Seven dimensions of learning (Ambrose et al., 2010)	Six aspects of class climate
Format	Flexible: structured items, open-ended questions, or both; faculty choose dimensions	Standardized: Likert-scale items and structured open-ended prompts
Anonymity	Anonymous	Anonymous
Personnel file	No	Yes
Required	No (instructor's choice)	Yes (institutional requirement)

Dimension	Formative Learning Feedback	Summative SPLE
Input for self-reflection and iterative growth	Yes	Yes

The formative feedback process exists because some of the most useful feedback students can provide — what is working, what is not, what they wish were different — is most valuable *before* the course ends, when the instructor can still act on it. The summative SPLE, by contrast, captures the student's experience of the full term and provides data for institutional evaluation. **These are complementary but distinct functions.** Both instruments provide input for self-reflection and iterative growth (Dimension 5 of the TEval framework) — the formative process by giving instructors actionable feedback while the course is still in progress, and the summative instrument by revealing patterns across terms that inform longer-term development.

3.2 Formative feedback as the home for open-ended questions

The SPLE committee voted to retain open-ended questions on the summative instrument under structured prompts and guardrails designed to minimize the equity bias that the literature documents in unstructured responses (see the SPLE proposal for the full evidence base). Open-ended questions also play a central role in the formative feedback process. In the formative context, results go only to the instructor, so potentially biased comments cannot influence personnel decisions. In a formative context:

- Results go only to the instructor, so biased comments cannot influence personnel decisions
- The instructor can contextualize comments with their knowledge of the class
- The developmental framing encourages constructive rather than evaluative responses
- There is no need to standardize or compare across instructors

Open-ended questions serve their intended purpose — giving students a voice and giving instructors actionable information — without the risks that attend their inclusion in the personnel file.

3.3 Optional structured check-in on SPLE dimensions

Instructors may optionally include a brief structured component — a "temperature read" on the six aspects of the SPLE — to get a quick snapshot alongside the formative items. This is **not** the SPLE itself; it is a lightweight check-in that uses the same conceptual dimensions.

Optional: Quick check-in on learning environment dimensions

Please indicate how you have experienced each of the following so far in this course.
Response options: **Positive experience**, **Mixed experience**, **Negative experience**, **Not sure / Not applicable**

Dimension	
Regard for Students — feeling treated with regard	○ Positive ○ Mixed ○ Negative ○ N/A
Consistent Communication and Enforcement of Expectations — feeling that all students are treated equitably	○ Positive ○ Mixed ○ Negative ○ N/A
Perceived Course Coherence — seeing how course elements connect	○ Positive ○ Mixed ○ Negative ○ N/A
Participatory Climate — feeling comfortable asking questions and sharing ideas	○ Positive ○ Mixed ○ Negative ○ N/A
Access to Instructor and Instructor Resources — feeling able to access help when needed	○ Positive ○ Mixed ○ Negative ○ N/A
Responsive Learning Environment — feeling that the learning environment is responsive to all students	○ Positive ○ Mixed ○ Negative ○ N/A

This structured component serves two purposes: it gives the instructor an at-a-glance summary of the dimensions they will be evaluated on at the end of the term, and it helps students become familiar with the conceptual framework before they encounter the summative SPLE. It is **not** scored, reported, or retained beyond the instructor's own use.

3.4 What the Formative Learning Feedback is *not*

To prevent misunderstanding, the following points should be communicated clearly to both instructors and students:

- The formative feedback process is **not** the Student Perceptions of Learning Experience instrument. The SPLE is the summative instrument administered at the end of the term.
- The formative feedback process is **not** part of the personnel file. No administrator, department chair, or personnel committee will see the results.
- The formative feedback process is **not** required. It is a tool available to instructors who want real-time feedback on the learning environment in their course.
- The formative feedback process is **not** anonymous feedback about the instructor's

teaching ability. It is anonymous feedback about the student's learning experience — the same conceptual framing as the SPLE, but in a developmental rather than evaluative context.

- The formative feedback process **does not** replace the SPLE. Students will still complete the summative instrument at the end of the term regardless of whether formative feedback was collected.

Chapter 4

Administration and Closing the Loop

4.1 Three touchpoints model

The formative feedback process can be used at **any point in the term** — once, twice, or at multiple touchpoints. Different dimensions of learning are most actionable at different stages, and the framework below suggests which dimensions fit naturally at each stage.

Table 4.1: Recommended dimensions by touchpoint

Touchpoint	Timing	Recommended Dimensions	Rationale
Early	Weeks 2–3	Prior Knowledge, Knowledge Organization	Students are orienting to the course; feedback on whether the course is connecting to what they know and how it is organized is most actionable here
Mid	Weeks 6–8	Motivation, Practice & Feedback, Student Development & Course Climate	Students have enough experience to report on engagement, the quality of practice and feedback, and the classroom climate

Touchpoint	Timing	Recommended Dimensions	Rationale
Late	Weeks 11–12	Mastery, Self-Directed Learning	Students can reflect on skill development and metacognitive growth; there is still time for the instructor to adjust the final weeks

This is a framework, not a prescription. An instructor might use only the mid-term touchpoint. Another might use all three but select only one or two dimensions at each. The instrument is a menu: faculty choose the dimensions that match their course goals and the timing that makes sense for their schedule.

4.2 Mode

Two modes are recommended. Instructors should choose the one that best fits their course:

4.2.1 Option A: In-class administration (~5 minutes)

1. The instructor distributes a brief paper form or displays a link/QR code to an online form
2. Students complete the feedback anonymously during class time (approximately 5 minutes)
3. The instructor may remain in the room (unlike the summative SPLE, the formative feedback process is not an evaluative instrument and does not require the instructor to leave)
4. If using paper, the instructor collects the forms; if using an online form, responses are submitted digitally

4.2.2 Option B: Online administration (3-day window)

1. The instructor sends students a link to an anonymous online form (Google Forms, Qualtrics, or similar)
2. The form is open for 3 days
3. The instructor sends one reminder during the window
4. Responses are anonymous

In-class administration is preferred because it typically produces higher response rates and takes only a few minutes. However, online administration may be more practical for large-enrollment courses or courses with irregular meeting patterns.

The technology already exists

Several universities have deployed online platforms that allow instructors to build and deploy formative feedback surveys in minutes. UC Irvine's EEE Evaluations system, for example, lets instructors select dimensions, choose items, set a collection window, and view results — all through a self-service web interface. Implementing a similar tool at Cal Poly would not require building from scratch; the CTLT could adapt existing survey infrastructure to offer a comparable experience.

4.3 Anonymity

All responses must be anonymous. The feedback should not collect names, student IDs, or any other identifying information. If using an online platform, the instructor should verify that the platform's settings do not record respondent identities or email addresses.

4.4 Supplementary open-ended items

In addition to the structured items drawn from the seven dimensions, instructors may include open-ended questions at any touchpoint. The following three types of items are recommended as a starting set:

Item 1: What is working

An open-ended item asking students to identify aspects of the course that are helping their learning. This provides positive reinforcement and helps the instructor understand which practices to continue.

Item 2: What could change

An open-ended item asking for constructive suggestions about the learning experience. The framing should keep the focus on the student's experience rather than inviting a judgment of the instructor.

Item 3: Open channel

An open-ended item providing space for concerns that may not fit neatly into the first two — issues of climate, inclusivity, accessibility, or anything else the student wants the instructor to know.

These open-ended items serve their intended purpose — giving students a voice and giving instructors actionable information — in the formative context where results go only to the instructor. As discussed in Chapter 2, this is the appropriate home for unstructured

feedback.

4.5 Closing the loop

The most critical step in the formative feedback process is **closing the loop** — the instructor's public response to the feedback received. Without this step, the feedback is a data-collection exercise that may actually *reduce* student trust if students feel their feedback was ignored.

4.5.1 What "closing the loop" means

Within one week of collecting feedback, the instructor should dedicate **5–10 minutes of class time** to:

1. **Acknowledge the feedback.** Thank students for participating and confirm that the responses were read.
2. **Summarize the themes.** Identify 2–3 themes that emerged from the responses. Be specific: "Several of you mentioned that the pace of lectures is too fast" is more effective than "I got some feedback about the course."
3. **State what will change (if anything).** If the feedback points to a change the instructor is willing and able to make, say so: "Starting next week, I'm going to pause more often during lectures for questions."
4. **Acknowledge what will not change, and why.** This is as important as stating what will change. Students respect transparency about constraints. What erodes trust is silence. If the feedback points to something the instructor cannot or will not change, explain why: "A few of you asked for fewer assignments, but the assignment sequence is designed to build skills progressively, so I'm going to keep the current schedule. What I *can* do is provide clearer guidance on how to prioritize your time."

4.5.2 Template for closing the loop

Instructors may find the following template useful for structuring their in-class response:

Closing-the-loop template

Thank you for completing the feedback check-in. I read every response. Here is what I heard and how I plan to respond.

What's working well:

- [Theme 1, in the students' words]
- [Theme 2]

What you'd like to see changed:

- [Theme 1]: Here is what I plan to do about this: [specific action]
- [Theme 2]: I understand this concern. Here is why the current approach is set

up this way: [brief explanation]. What I *can* adjust is: [specific action, if any]

Other concerns raised:

- [If applicable, address any climate or environment concerns with care and specificity]

I appreciate your willingness to share your experience. If you have follow-up thoughts, my office hours are [time/place] and you can always reach me at [email].

4.6 In sum

The **Formative Learning Feedback** process complements the summative **Student Perceptions of Learning Experience** by covering the dimensions of effective teaching that students can observe and report on but that fall outside what they can validly evaluate for personnel purposes. It is voluntary, developmental, and shared only with the instructor — making it the appropriate home for both structured feedback on teaching practices and the open-ended questions that the literature identifies as too susceptible to bias for inclusion in a personnel file. The seven dimensions, drawn from Ambrose et al. (2010), give faculty a research-grounded menu of feedback options they can deploy at any point in the term, on their own terms, in service of their own growth as educators.

4.7 Bibliography

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., and Norman, M. K. (2010). *How Learning Works: Seven Research-Based Principles for Smart Teaching*. San Francisco: Jossey-Bass.

Angelo, T. A., and Cross, K. P. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers* (2nd ed.). San Francisco: Jossey-Bass.

Austin, A. E., Finkelstein, N. D., Greenhoot, A. F., Ward, D., and Weaver, G. C. (2025). *Transforming College Teaching Evaluation: A Framework for Advancing Instructional Excellence*. Cambridge, MA: Harvard Education Press.

Clark, D. J., and Redmond, M. V. (1982). *Small Group Instructional Diagnosis: Final Report*.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., and Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences*, 116(39), 19251–19257.

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

2026-05-07

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.