U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Kathleen:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **ANTHROPOLOGY**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E/D**	**E**	**E**	**E/D**	**E/D**
Primary	**E/D.** The PLO could be improved by specifying what some of the terms and concepts are. Several of the components of the PLO were not assessable because students "addressed topics that did not lend themselves to demonstration of proficiency."	**E.** The assessment plan includes only direct evidence. Results are biased by the fact that the work of several students who were unable to complete the assignment could not be included.	**E.** The reliability of scoring should probably be checked by including some degree of overlap in which multiple raters assess some of the papers. While this was done initially with a single paper, the purpose of that was primarily to develop the rubric; inter-rater reliability needs to be monitored in the actual operational assessment.	**E/D.** Clear data are missing for some aspects of the PLO for the reason described under "assessable PLO."	**E/D.** The recommendations focus unduly on possible changes in prerequisites.
Secondary	**E/D.** PLO is quite broad; could be broken into two or three PLOs. In some respects, the criteria for assessing the PLO could themselves serve as useful PLOs, as they're relatively more focused, clearer, descriptive, and measurable.	**E.** Insufficient sample size, and two essays per rater will not provide sufficient range for rater to consider and to apply the rubric.	**E.** There's need for more overt and active calibration among raters.	**E/D.** Thoughtful consideration of results—which should inform further assessment efforts. Still, it's difficult to generalize from such a small sample, and the criteria for assessment could be more specifically discussed.	**E/D.** Aside from extensive focus on prerequisite changes—which can distract from the assessment at hand—there are useful conclusions about how to streamline and focus assessment, and to actively promote student attainment of PLO.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The program has chosen to focus on direct evidence from a class that is required for all majors and minors, and the assignment used is one that has the potential to yield rich information about the PLO.

Secondary Reviewer: The program is attuned to the intricacies of student written expression as it pertains to the PLO.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: A process for ongoing monitoring of rater reliability should be implemented; something needs to be done to address the fact that some components of the PLO could not be assessed using the papers.

Secondary Reviewer: The program could follow through on some of its useful conclusions about student work to make the assessment process more complete and rigorous—including active calibration, qualification and quantification of the assessment exercise.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

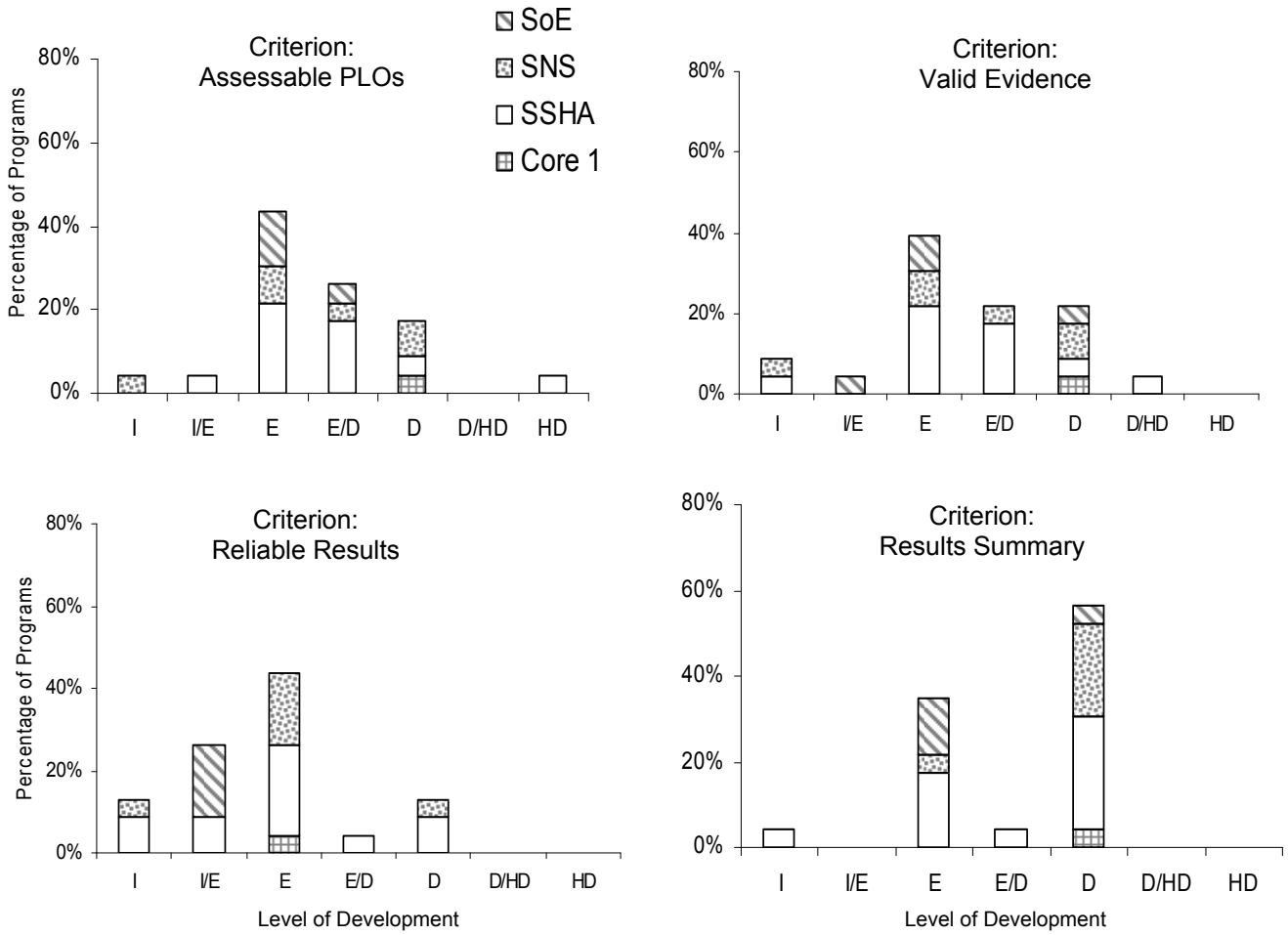
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



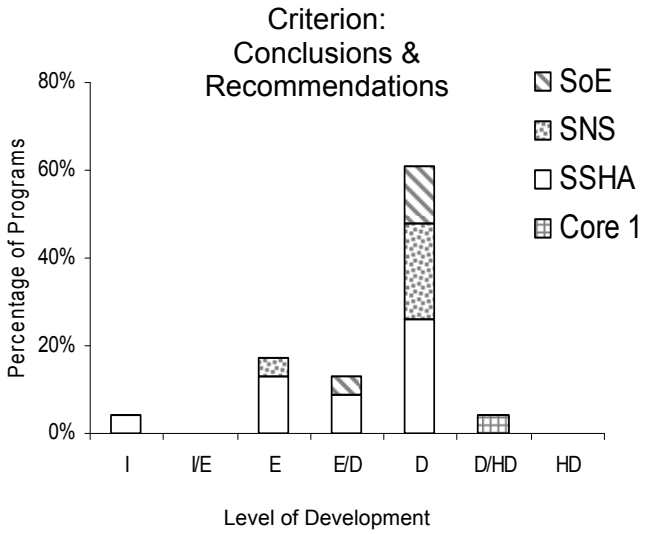


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

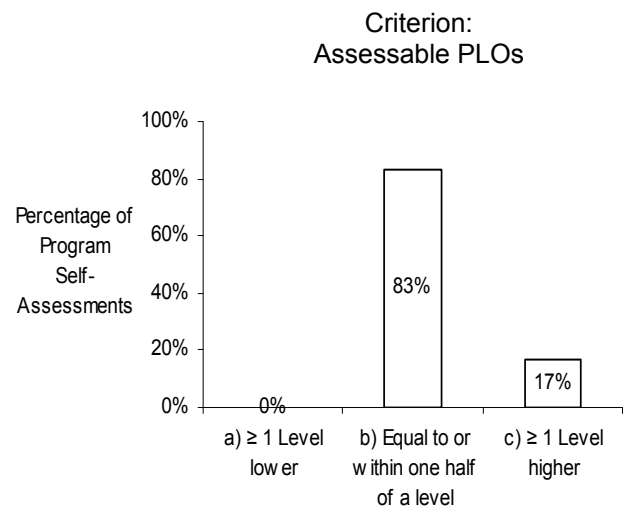
These results indicate that

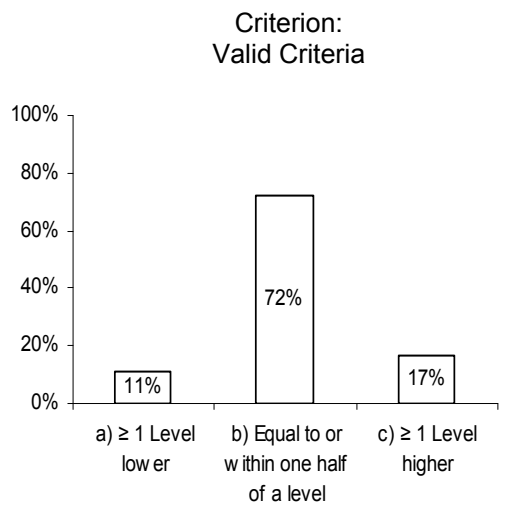
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

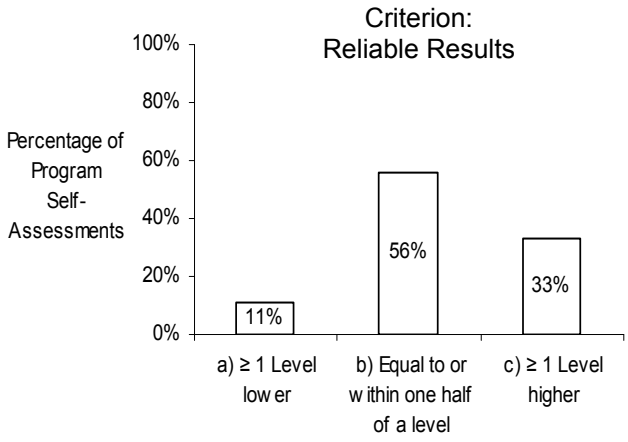
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

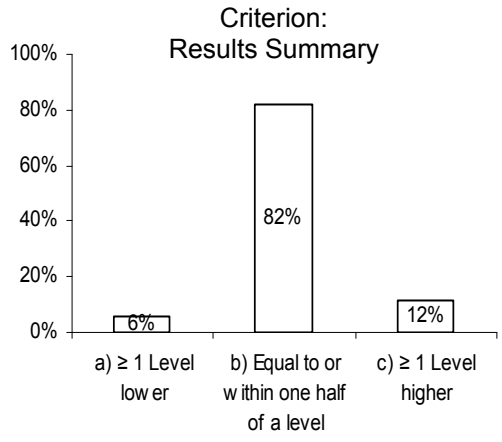
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









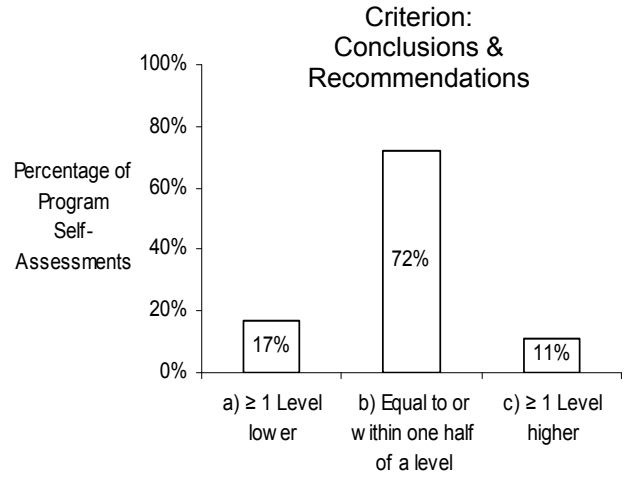


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Arnold:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **APPLIED MATH**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**E/D**	**E**	**D**	**D**
Primary	E/D – have a rubric and realize that it isn't up to the job and are considering adopting the one they use for grading homework which supports alignment of the curriculum. The holistic rubric doesn't appear specific to the PLO however – lacking criteria and standards that could support the development of faculty consensus and student awareness of key attributes.	E/D – closer to D as with the first criterion. Alignment between evidence and rubric is questionable as noted by the faculty. Clearly the faculty didn't feel the evidence was sufficient.	E – faculty did not calibrate but did examine inter-rater reliability.	D – provide results but wish they had also included summary of analyses they had performed.	D
Secondary	E – Need to have more articulate and clear assessment criteria. The rubric is still in early stages of development	E/D – I feel that faculty collect relevant and sufficient evidence for the outcomes. However, evidence needs to be assessed against new criteria to produce meaningful results.	E/D – Faculty routinely check for inter-rate reliability. However, I don't see any evidence of reviewers' calibration.	D – Results are clearly presented in the report.	D – This report clearly states conclusions, implications, and recommendations for improvement.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The interest and involvement of all faculty is an incredible strength as it is critical to development of shared expectations among faculty over time and, in turn, consistent support for development of student skills throughout the curriculum.

Secondary Reviewer: The strength lies in the great collaborations of faculty within the program.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) Continue to push to develop a shared understanding among faculty with respect to the skills and knowledge that underpin each PLO (the criteria by which student work is evaluated) and, for each PLO, articulate this in a rubric that can be used by faculty, TAs and perhaps eventually students. It seems clear the faculty realize their differences and the need for reconciliation in order to support student learning across the curriculum. While the holistic rubric has the advantage of being something that is used regularly in courses and thus promotes alignment between program assessment criteria and scoring practices, it doesn't seem to identify skills specific to the PLO and in that way doesn't promote the development of consensus. (2) Related to the above, it is obviously important for the assessment questions to allow students to demonstrate the expectations outlined by the rubric. It is equally important that the curriculum support student development of expectations in the rubric. Maybe one way to help promote this alignment and to identify shared expectations among faculty is to ask faculty across the program for examples of homework and exam questions they really value as reflective of the kind of learning they want students to demonstrate with respect to a given PLO. A discussion of the similarities and differences and strengths and weaknesses could be a starting point for identifying a shared set of criteria and expectations for students proficiencies at given points in the program.

Secondary Reviewer: Agree with Primary Reviewer.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

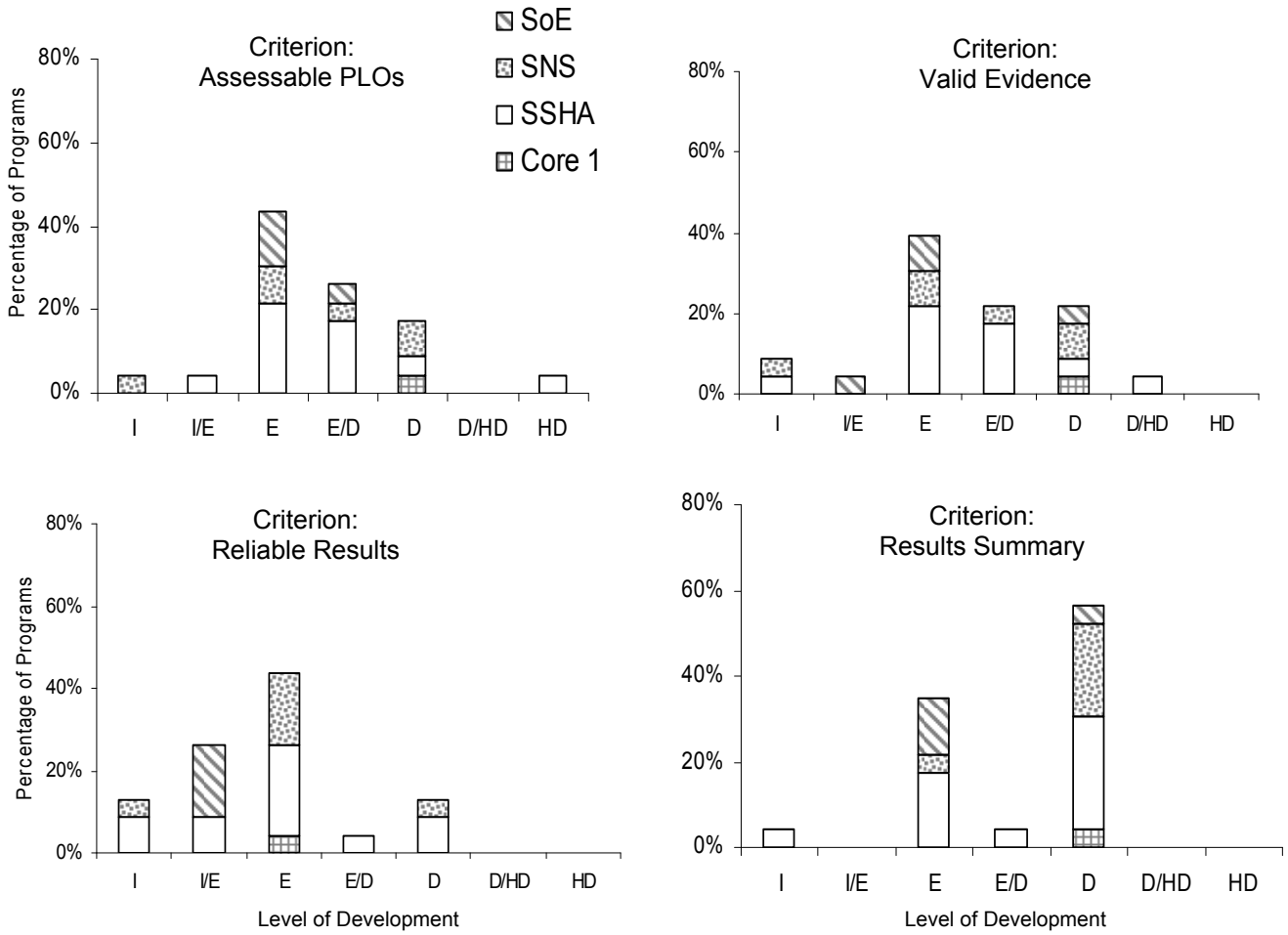
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



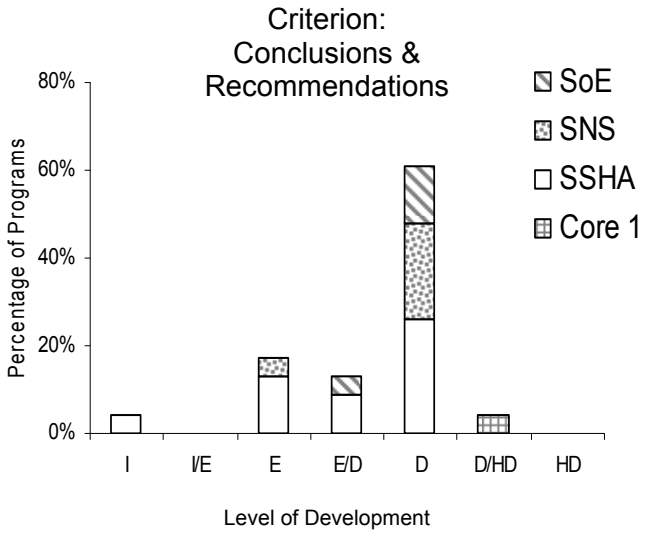


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

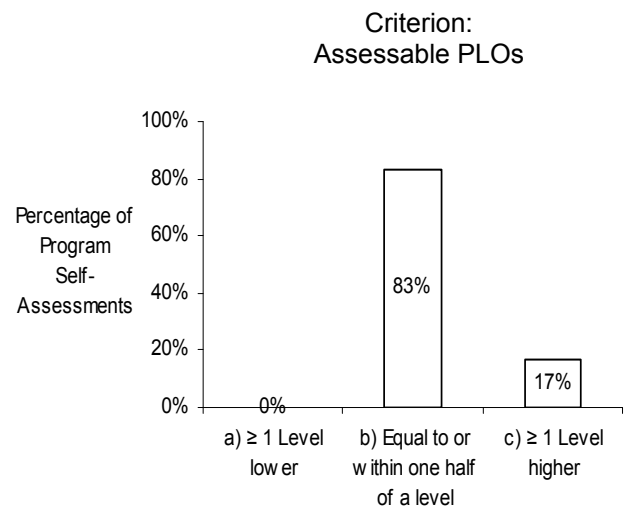
These results indicate that

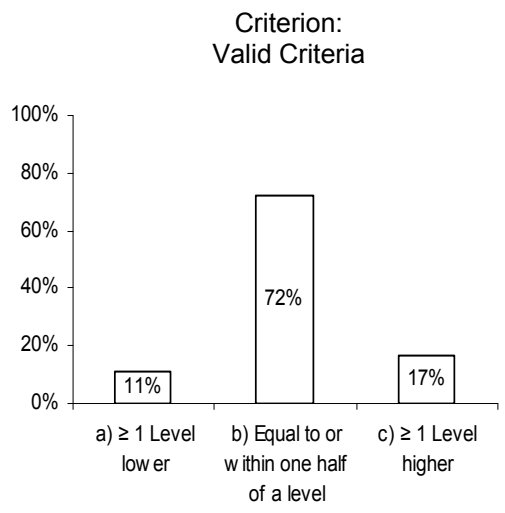
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

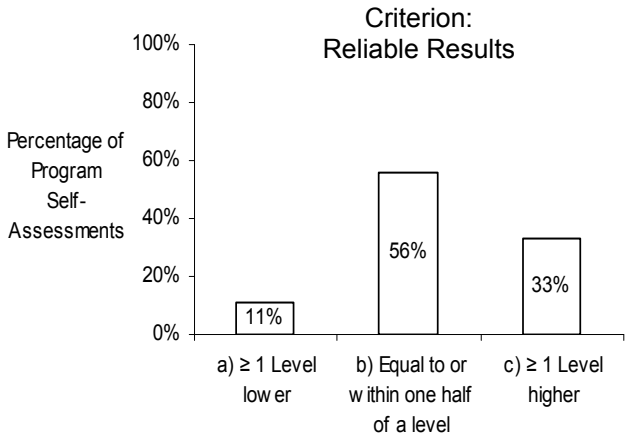
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

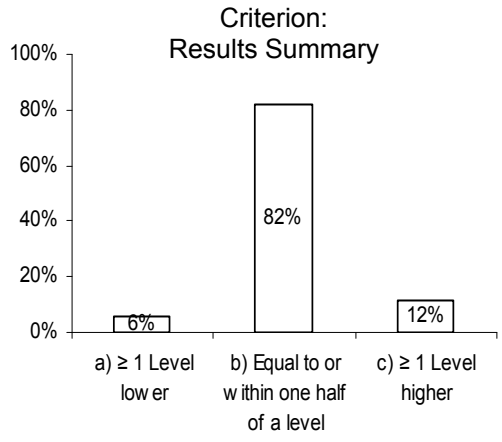
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









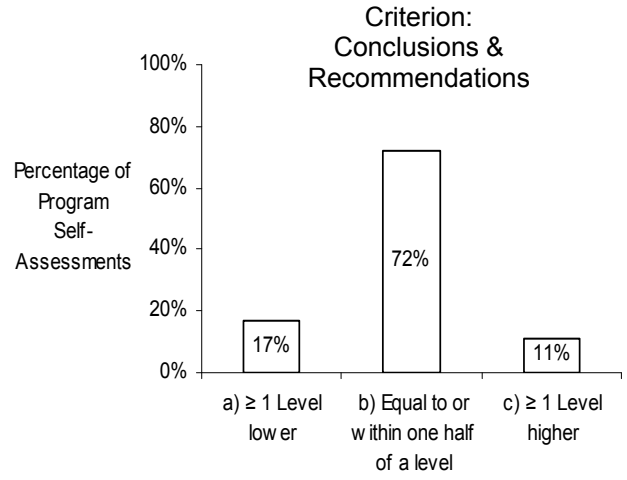


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Susan:

Please find attached feedback regarding the Arts' 2009 PLO Assessment Reports (MAP and GASP), provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **GASP**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**I**	**I**	**E**	**I**
Primary	**E** – PLO is assessable. No assessment criteria are shared.	**I** – the evidence provided is in the form of the mean final grade. It is impossible to tell if the evidence is aligned; a single professor assessed this work and there is no discussion of input from other faculty, personalized not programmatic criteria were used to judge student performance	**I** – only a single reviewer, no assessment criteria beyond final grades provided	**E** – results are included but only in the form of the mean final grade, key data missing that would show alignment between assignments and PLO and assessment criteria	**I**
Secondary	**E** – Although PLO is assessable, what is meant by "technical" or "theoretical" could be clarified.	**I** – No data save for mean grade; no discussion of assessment process. Not clear how assignments are constructed.	**I** – No assessment (or grading) rubric; no elaboration of assessment criteria.	**I/E** – Difficult to tell how students attained results, as only reported data is grade.	**I** – No discussion of implications or future assessment efforts (save for mention of resource need).

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The program recognizes that its assessment efforts are limited to a single course and are not programmatic in nature, providing an opportunity to expand the assessment work to examine student performance as a result of participation in the program in the next round of assessment.

Secondary Reviewer: The program is aware that it can build out its assessment efforts beyond a single course (so as to develop program definition), and notes that such efforts can be based on critical analysis of specific subject matter from diverse sources.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) GASP needs to move toward a programmatic approach to assessment that involves more than one faculty member applying an explicit set of criteria and standards (as articulated in a rubric) to specific examples of student work selected for their relevance to demonstrating the proficiencies articulated by the PLO. Unless a specific rubric is generated to describe the relationship between a grade and performance on a given exercise, grades do not help communicate to other faculty what attributes the grading faculty was using to assess student learning, rendering evaluation criteria personal. (2) Similarly, to assess student learning from a programmatic perspective GASP needs to identify student work from more than one class to ask if the curriculum/program is cultivating the desired abilities at the desired levels. (3) These assignments should be designed specifically to elicit responses that will allow faculty to evaluate student performance with respect to the PLO and the criteria and standards that have been elaborated.

Secondary Reviewer: Overreliance on grades—without explanation of how they were awarded nor inclusion of a rubric of how students attained them—can take assessors out of the assessment equation, and diminish active student-learning-oriented results. Report should describe the assignments students completed, demonstrate how students performed on them, show how raters assessed them, and include examples of student work. Also, there can be more consideration of how the assessment effort maps onto programmatic definition.

PLO Report Review Summary

Name of Program: **MAP**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**E**	**E**	**E**	**E**
Primary	E – the PLO needs some refinement in terms of describing how the students will demonstrate their "understanding". Based on the assessments reported, it looks like this might be not through the production of art but through written forms.	E – it's clear that direct and indirect forms of evidence are being collected in various courses. However, the assessment is not focused on the minor but on those enrolled I classes so there isn't a sense of what learning looks like as a result of the program	I/E – Although faculty appear to be applying individual criteria to the review of student work; the review and summary work by the FAO helps to cultivate some consistency across the program	E	E – recognized some of the constraints of the method employed. Changes will happen.
Secondary	E	E	E	E	E

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The breadth of faculty involvement, the repeated nature of the assessment, the amount of student knowledge considered, and the implications of this effort for the entire curriculum are remarkable aspects and strengths of MAP's assessment effort.

Secondary Reviewer: I agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) MAP has put tremendous energy into the assessment of the PLO within each of course, resulting in important course-level reflections and feedback. The amount of work could be substantially reduced, however, if the program's assessment work was focused instead on the minor and, thus, on the subset of students who are pursuing a Minor in Arts.

(2) Assessment that is focused on students completing the minor will provide faculty with the opportunity to elaborate for each PLO descriptions of expected student performance that reflect the cumulative impact of the curriculum of the minor, rather than of a single course.

For example, in relation to the first PLO which asks students to "understand the principle attributes and mechanics of art techniques," the faculty would determine how specifically students will demonstrate "understanding" at the time they complete the minor. This can be done in various ways including, for example, in the form of a written exercise (as has been done in some of the courses) or in the actual production of a piece of art that demonstrates awareness of key attributes and mechanics for a given medium. For this form of assessment, the faculty would then elaborate a rubric defining the set of qualities that describe what they would look for in the piece of student work to demonstrate "principle attributes and mechanics of art techniques", and what each of these criteria look like for students of differing skill levels.

By establishing shared goals for the graduates of a program in this way, the faculty can then reflect on the contributions their own courses make to these desired proficiencies. CRTE staff are available to support rubric development and information about rubrics can be found on the CRTE website - http://crte.ucmerced.edu/assessment-guide#rubric.

(3) With these shared expectations more precisely defined, the faculty can then also revise the PLO, replacing "understand' with a description that shows how learning will be applied. For example, the PLO might read "students will be able to produce a piece of art that illustrates the principle attributes and mechanics of a preferred art technique."

Secondary Reviewer: While I agree (especially with point 3), I can see why, given that the point of the program is to serve the UCM population broadly, rather than minor narrowly, that the FAO has chosen to focus on the course level. It's something of a catch-22: shift the focus of assessment and risk losing the program's real goal, or keep the current breadth and die trying to assess everything. Perhaps rolling assessment, that is using a sample of courses each year, and roll through all of the courses in a cycle?

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

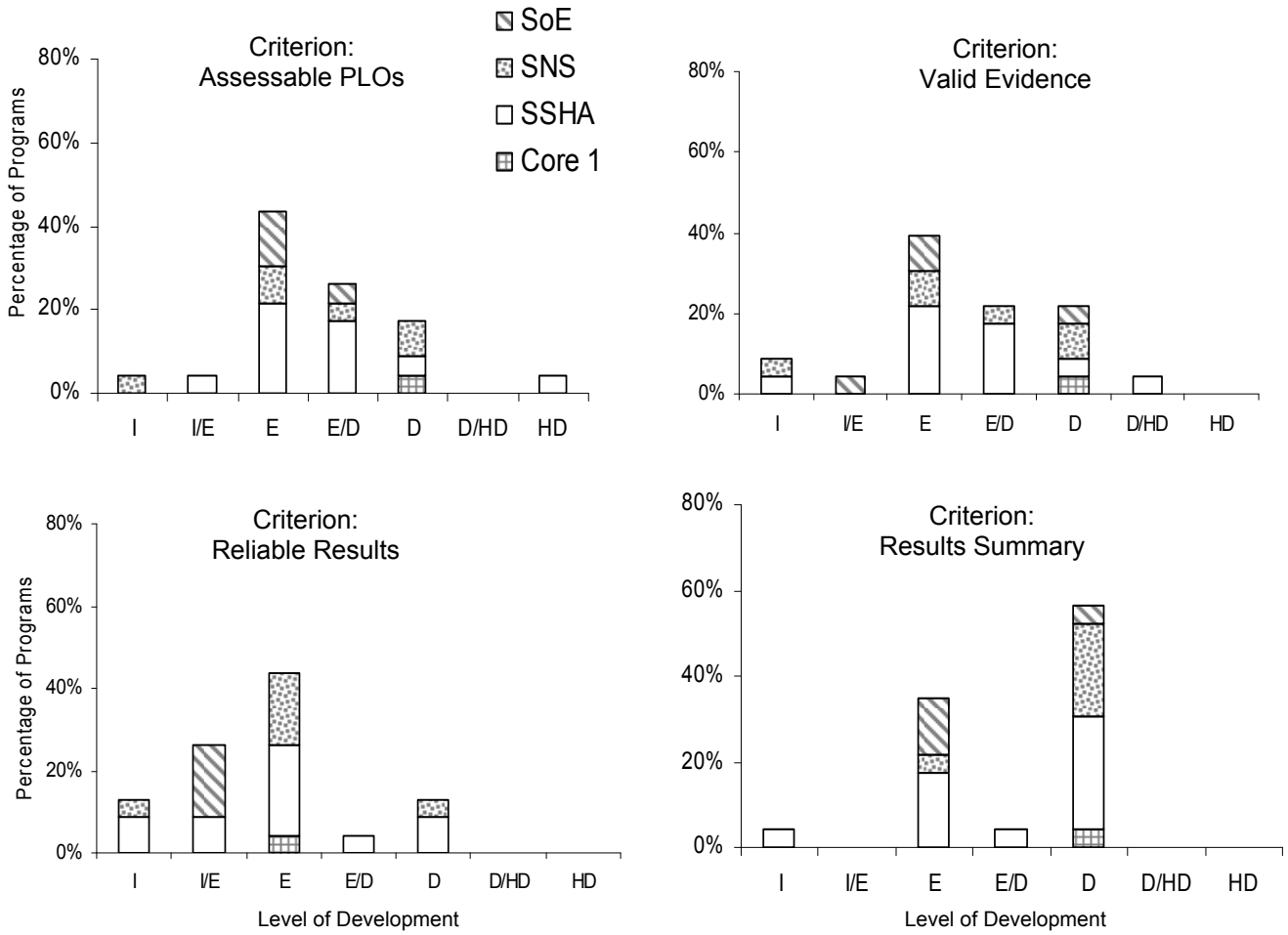
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



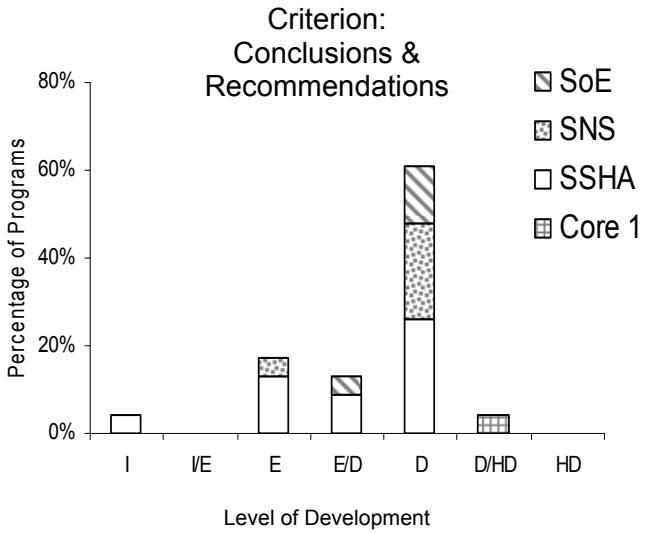


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

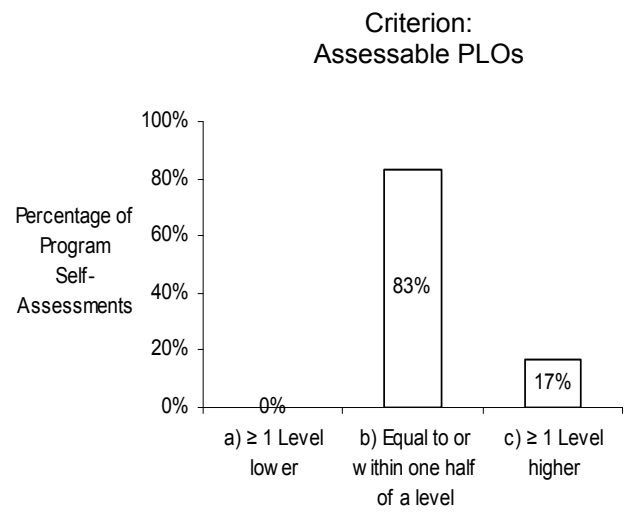
These results indicate that

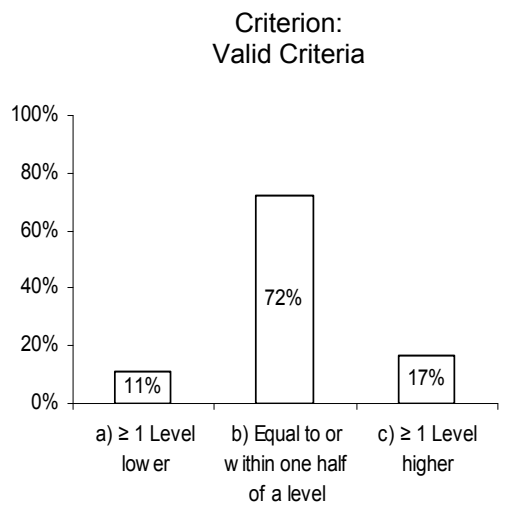
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

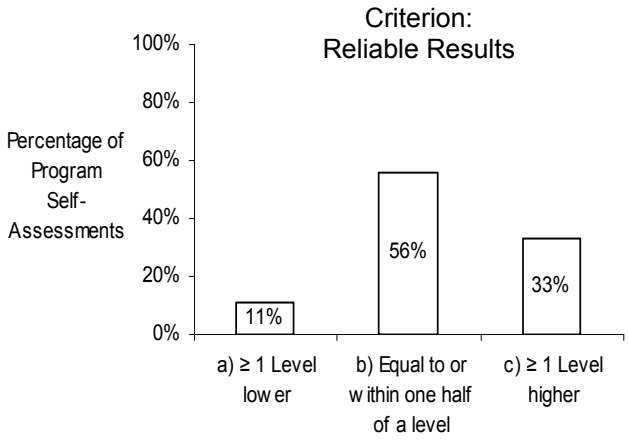
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

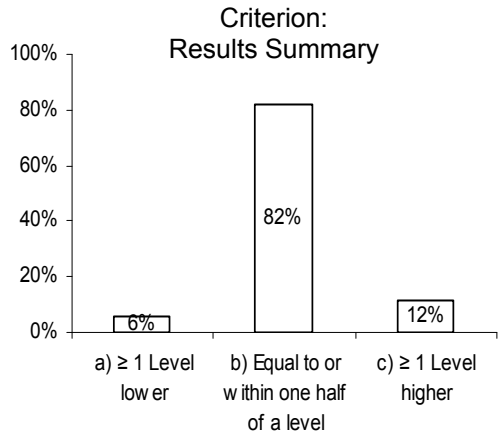
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









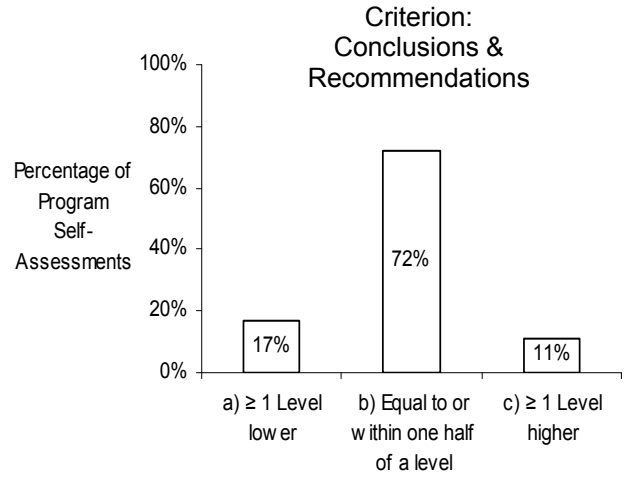


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Patti:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **BIOLOGY**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average					
Primary	D – have reasonably assessable outcomes. The rubric is defined, but requires more development.	E – Faculty have agreed on the evidence to be collected for the PLO. However, only direct methods to assess the PLO were included. Some evidence could not be used for the assessment.	E/D – Four reviewers clearly applied assessment criteria and they checked for inter-rater reliability. However, some direct evidence was not assessed.	E/D – Presented results delineated some aspects of the PLO. However, missing data for other aspects of the PLO.	D – all elements were described. Report contains clear conclusions, implications, and suggestions.
Secondary	E – it is not clear precisely what intellectual skills the students should exhibit as the verb "understanding" is vague. For example, will students be expected to demonstrate their understanding in the form of explanations, analysis, synthesis, the ability to predict? The PLO needs to provide more guidance to all stakeholders.	Agree	D –clearly need to work to improve inter-rater reliability, which could benefit from refining the rubric.	D – had one line of evidence and fully summarized the results, including evaluating them against a programmatic goal for student proficiency levels	E/D – questions of reliability are not examined

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A real strength of the program's assessment efforts is that the number of faculty involved in crafting the rubric and participating in the assessment work.

Secondary Reviewer: I would add that the faculty undertook a programmatic approach to assessment as exemplified by 1) the programmatic goal for student proficiency against which they evaluated student performance and 2) the fact that they intended to examine evidence of student learning related to this PLO from multiple courses.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) Continue to push to develop multiple, complementary forms of evidence to assess students' knowledge and skills in relation to the PLO. It is clear that faculty realize that the rubric could not be applied to certain embedded questions. Do the current assessment criteria articulate the goal outcomes of the PLO? In the future, it will be important to develop assessment questions allow students to demonstrate the expectations outlined in the rubric. Consider discussing the assessment of "composition" skills with a writing program or CRTE consultant. (2) It will be important to gather indirect evidence to provide insight into the student's perceptions of their learning relative to the PLO. Like Anthropology, consider developing a set of questions that are administered to students as part of the graduating senior survey administered annually by IPA.

Secondary Reviewer: I would rate the following as the most important work of the program but agree the program also needs to push to develop a complementary, indirect source of evidence as well.

(3) Revise the PLO to describe specifically how students will demonstrate their understanding of the tenets of modern biology and the integration of cellular functions through the hierarchy of biological organization would help all stakeholders recognize the education intentions of this PLO. This is critical for faculty - they need to know what skill(s) they should support the development of through their individual courses - and for students who need to know what skills they should be practicing. It would also help the faculty to develop appropriate assessment tools, including as relevant a rubric with criteria more specific to the PLO, which is a need noted in the report. As an example, should the faculty be teaching students to

- *identify* examples of this kind of knowledge, or
- *apply* this knowledge to identify potential solutions to novel problems or
- be able to *make predictions* based on this knowledge, or
- *develop an hypothesis* and *design* an experiment involving the use of this knowledge?

To put it anther way, what it is the faculty want their graduates to be able to do with the knowledge described by this PLO when they join the workforce or matriculate in medical or graduate school? What are they preparing their students for? Related to this, the students would benefit most if the assessment tool asked students to engage in one or

more tasks that are more similar to those they would be expected to practice as graduates. Embedded questions could still be used but the intellectual activities/skills in which they engage the students could be more relevant to desired workforce or post-graduate skills. (Examples include the last three bullet points above.) This later point might be a longer term goal because these kinds of assessments would have to be implemented across the program in order to allow students to develop these abilities overtime with feedback from faculty and TAs. The program will not be successful if the only time the students meet a question related to the skills articulated in the PLO is in the form of the PLO assessment task. Finally, it would be helpful to revise the PLO to make it clearer. At the moment the PLO seems to describe either two separate sets of knowledge (the modern tenets and the ability to see how cellular functions influence the properties of higher levels of biological organization) or a single tenet articulated by the second phrase of the outcome.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

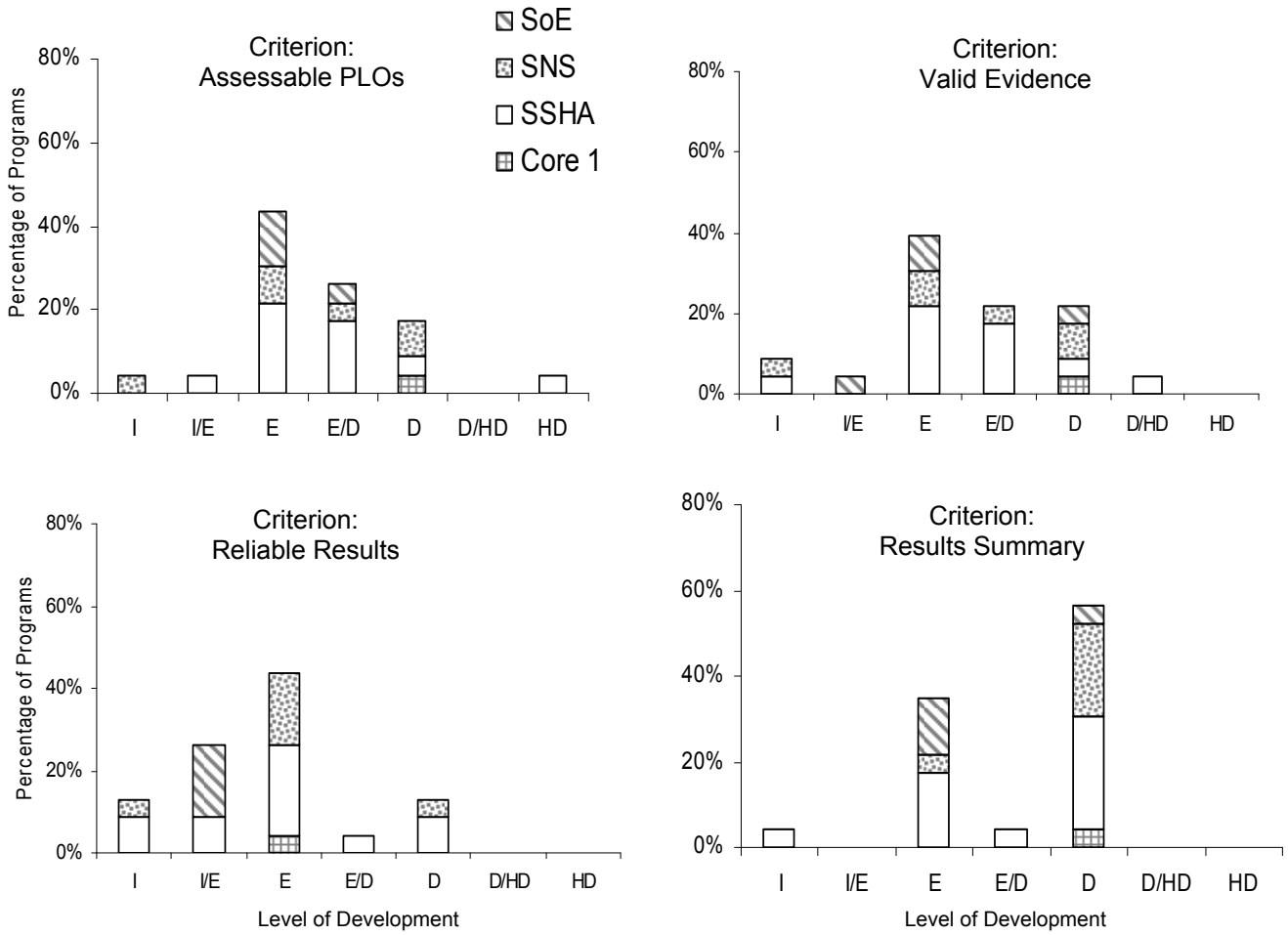
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



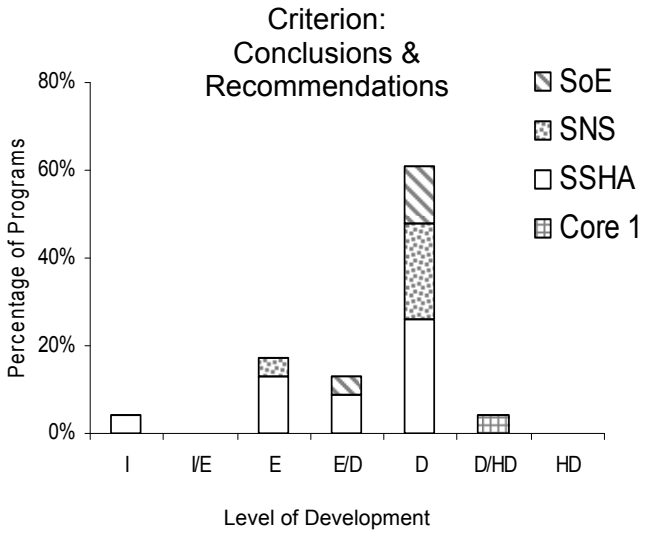


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

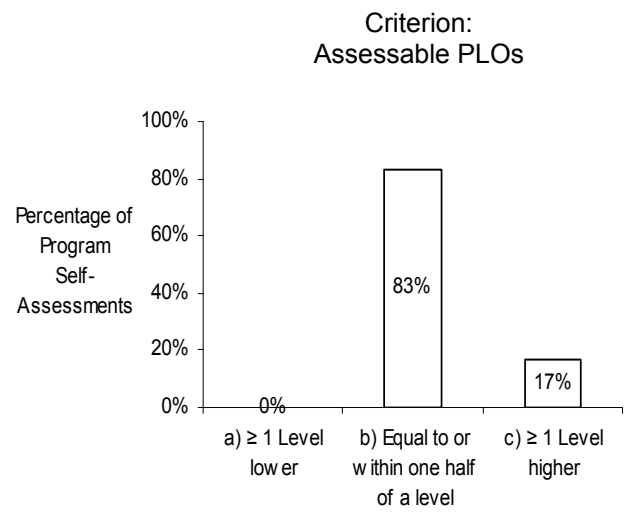
These results indicate that

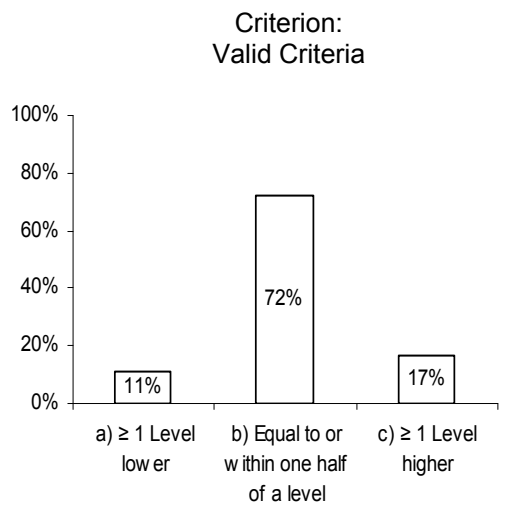
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

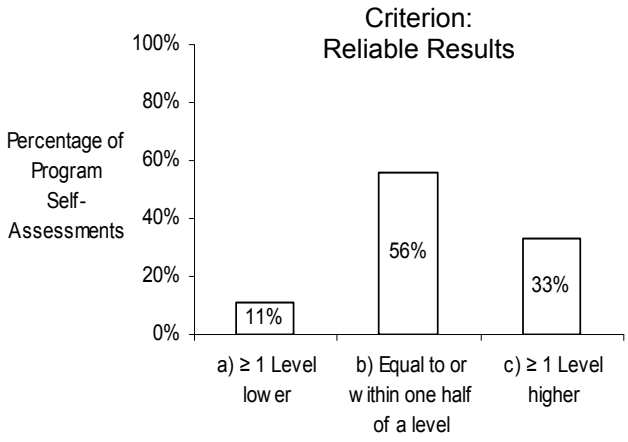
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

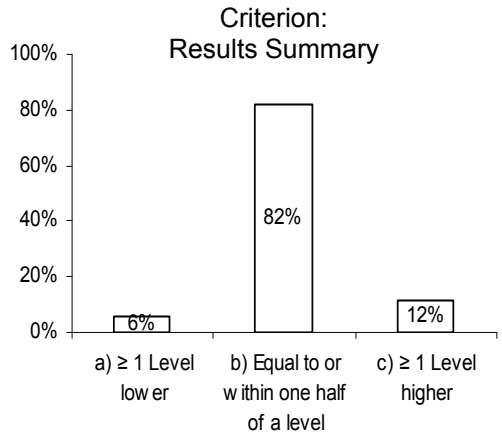
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









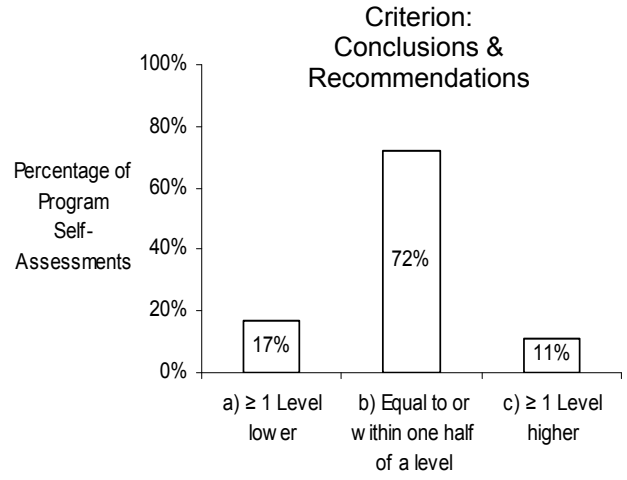


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Anne:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **CHEMISTRY**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**E**	**E**	**D**	**D**
Primary	E/D – PLO clearly assessable with specific verbs but lack any kind of shared criteria and standards	I/E – faculty are using personal criteria, not programmatic and so alignment between PLO, assessment tool and evaluation criteria problematic	E – faculty did examine inter-rater reliability but did not calibrate.	D – clear, meaningful summary of results,	D – have met all elements including recognizing limitations
Secondary	E – Neither clear assessment criteria nor rubric.	E – Evidence need to be assessed against clear criteria.	E – Not clear about reviewers' calibration before reviewing evidence.	D – Results clearly delineated for the evidence in summary format.	D – Have clear articulate conclusions, implications and recommendations for improvement.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A clear strength is the involvement of all faculty in the assessment work, which provides the opportunity for the development of consensus regarding the attributes student work should exhibit with respect to each PLO (i.e. criteria that define what looking for in student work) and associated levels of performance.

Secondary Reviewer: Agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer:(1) The assessment work is "programmatic" in focus in that work from diverse courses is examined but it is not "programmatic" in terms of working toward increasing faculty consensus about what student proficiency will look like at graduation. As an example, assessment of student lab skills is to be conducted in two different courses. However, two different instructors are slated to do this without, it appears, any rubric or other tool for elaborating shared criteria will be used to determine whether a student can "perform a chemical synthesis" or "operate equipment" in the separate classrooms. Similarly, in the recent assessment efforts each faculty member applied personal criteria to the assessment of the work they examined, without it appears, any discussion of what each individual was looking for in the

work. As a result, two faculty may have rated a paper as fair but it could be for two entirely different reasons.

To have truly programmatic assessment, chemistry needs to work to develop shared expectations for student performances, defining specifically what attributes of student work (or skills performance) will be examined and what defines "good enough" with respect to these criteria. One over-riding goal of assessment is to develop over time a shared, program perspective on what criteria are used in judging student work and what distinguishes different levels of proficiency with respect to each PLO(or relevant subsection). Once common intentions are established, faculty across the curriculum can help to develop student skills toward this common goal. Students also benefit because the curriculum is more unified.

(2) Related to 1, it would be helpful if in the future chemistry included in its reports examples of the questions that were assessed. This would permit programs to keep a record of what sort of work was done and how effectively in order to improve assessment. For example, it would be possible to evaluate alignment between the questions or prompts that were used to elicit student performance and the PLO.

Secondary Reviewer: Agree.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

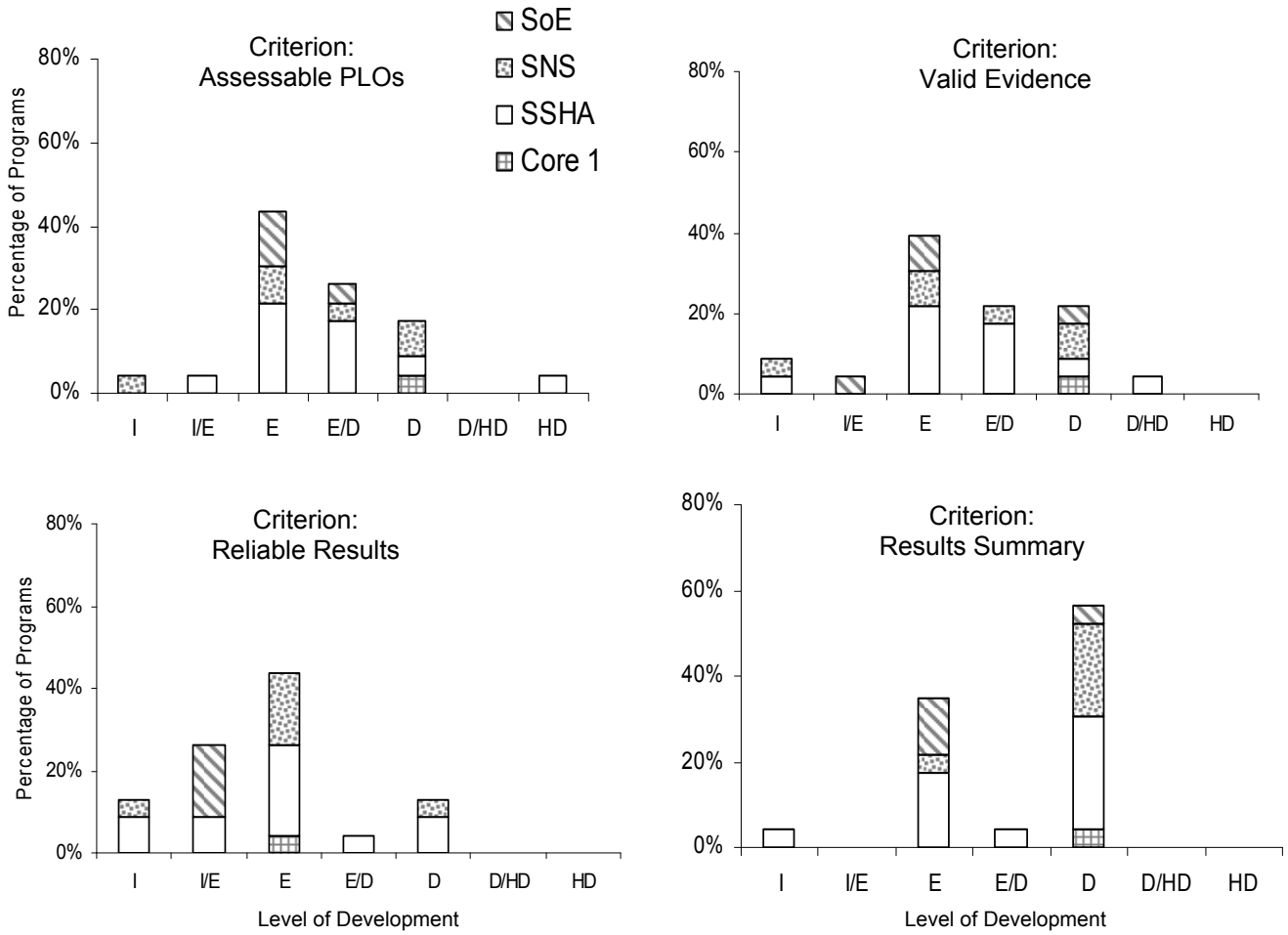
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



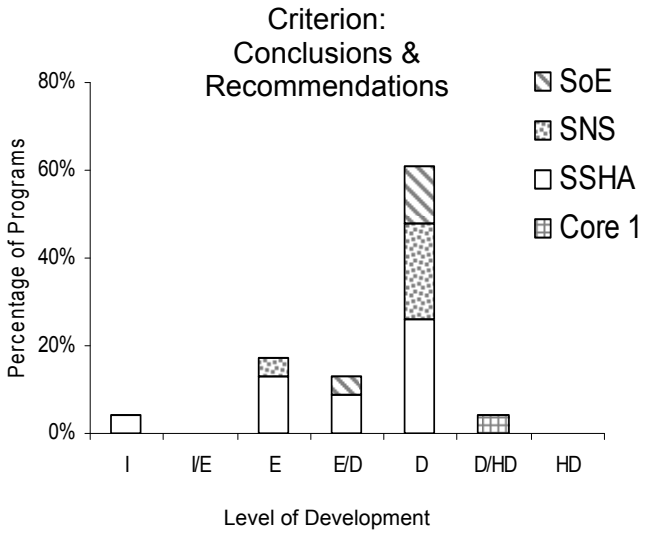


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

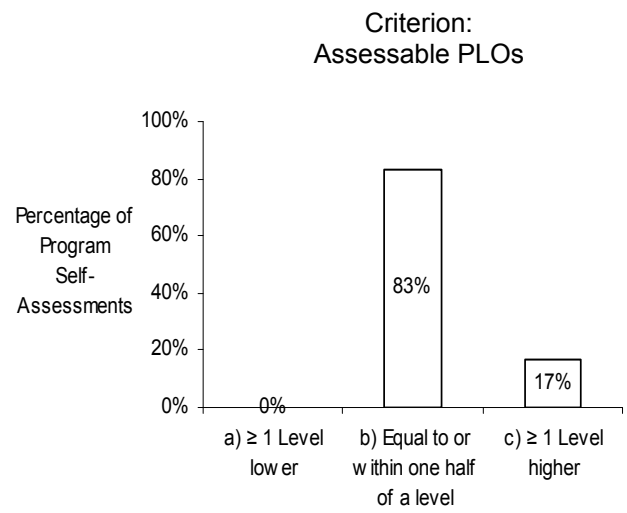
These results indicate that

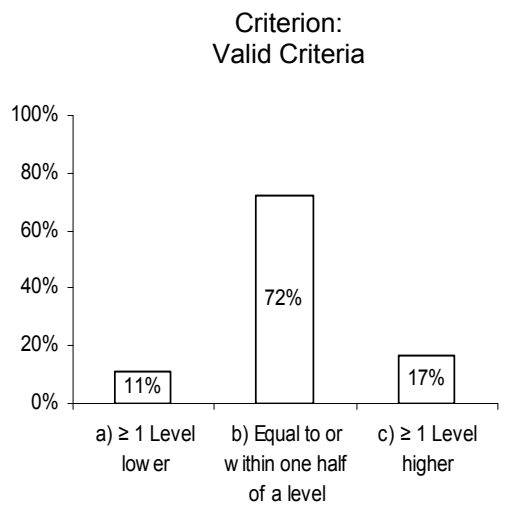
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

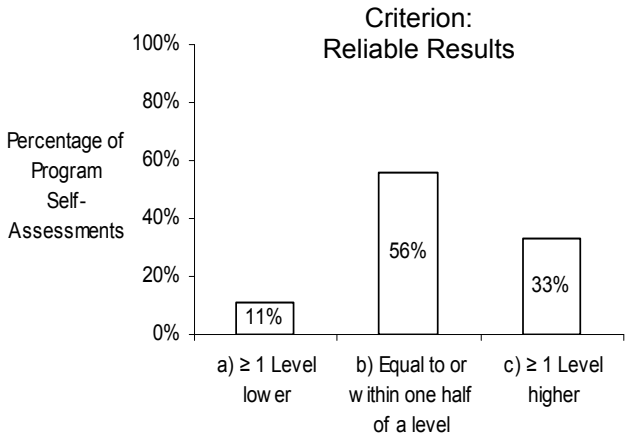
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

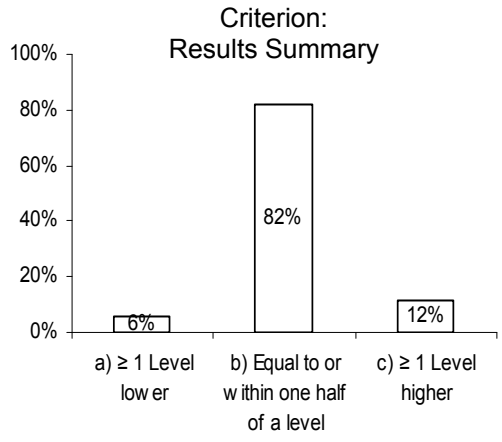
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









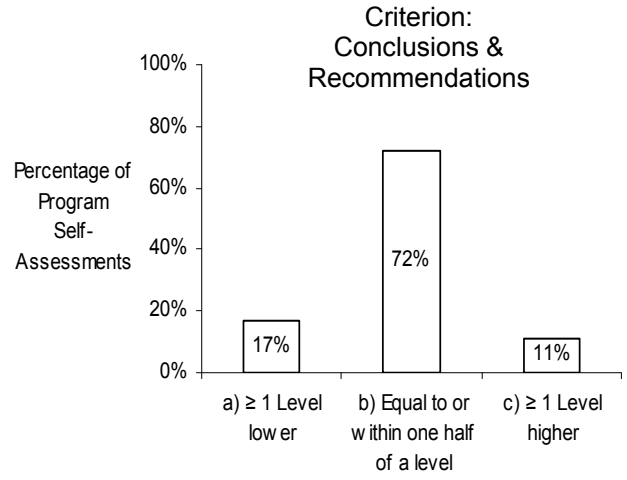


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Jeff:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **COGNITIVE SCIENCES**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**E/D**	**D**	**E**	**E**
Primary	E	E/D	D	E/D	E
Secondary	E/D	E/D The assessment does not consider indirect evidence.	D	E A tabular summary that shows counts of students in each category would be more useful than mean ratings for a rating scale that is ordinal level like this one. A sequence of three two-way cross-tabulations would also be useful. Completely redact participant names.	E

2) A clear strength of the program's assessment efforts.

Primary Reviewer: Inter-rater reliability is clearly demonstrated with concrete examples of student work that has been rated. However, if changes to the rubric for writing assessment suggested elsewhere in this document are implemented, the reliability for rating student writing (which was the highest) will undoubtedly suffer.

Secondary Reviewer: Agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: Strengthen the assessment criteria by (1) clarifying levels of achievement and by (2) focusing on more important aspects of student performance. (1) In distinguishing four levels of ability, one assumes equal gaps between the levels, yet the descriptions suggests otherwise. Distinctions particularly between levels 2 & 3, are vague in principle ("some degree" compared to "limited degree") though the rubric does lend clarity. Nonetheless, the gap between "some degree" and "full mastery" seems inordinately large, especially considering that "full mastery" regarding the "ability to write clearly and scientifically" is described as "similar in quality to what one find [sic] in a published study." Is this an appropriate standard for a soph/jr level class? In the examples rated #1, the writing clearly did not meet this standard, though it does seem to be of high quality for the grade level. Moreover, this standard raises real concerns in that the rubric emphasizes grammatical correctness above all. Professional standards in writing instruction usually privilege argument over correctness and ask instructors to consider grammatical correctness in the context of the conventions of a discourse community. Grammatical correctness should not be the single, or even the most important, measure of the "ability to write clearly."

Secondary Reviewer: I disagree with the assumption by the primary reviewer that a four-point rating scale implies equal gaps between levels. This is actually rarely true, which is why models exist that assess the probability of crossing boundaries between levels. However, the lack of an interval scale also implies that the mean ratings that are reported are not relevant.

Although grammatical correctness is important in scientific writing (which to me would imply that the conventions of the discourse community *should* involve careful attention to grammar), I agree that it would be good to place more emphasis on organization and clear development of arguments. Unfortunately, the high reliability of the writing ratings would undoubtedly suffer.

Synthesized Recommendations:

1) Quality similar to a published study is probably too high a standard for undergraduate program.
2) The rubric for writing assessment should focus more on argument development and less on grammatical correctness (although grammatical correctness can appropriately be *part* of what is assessed).
3) The anchor labels on the rating scale are vague, it would be desirable to attempt to space the achievement levels more evenly, and "full mastery" is probably too high a standard.
4) It would be methodologically more appropriate and more informative to the faculty to report the results of the assessment as frequencies of observations for each criterion and level of performance (or relative frequencies), for example in a tabular format.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

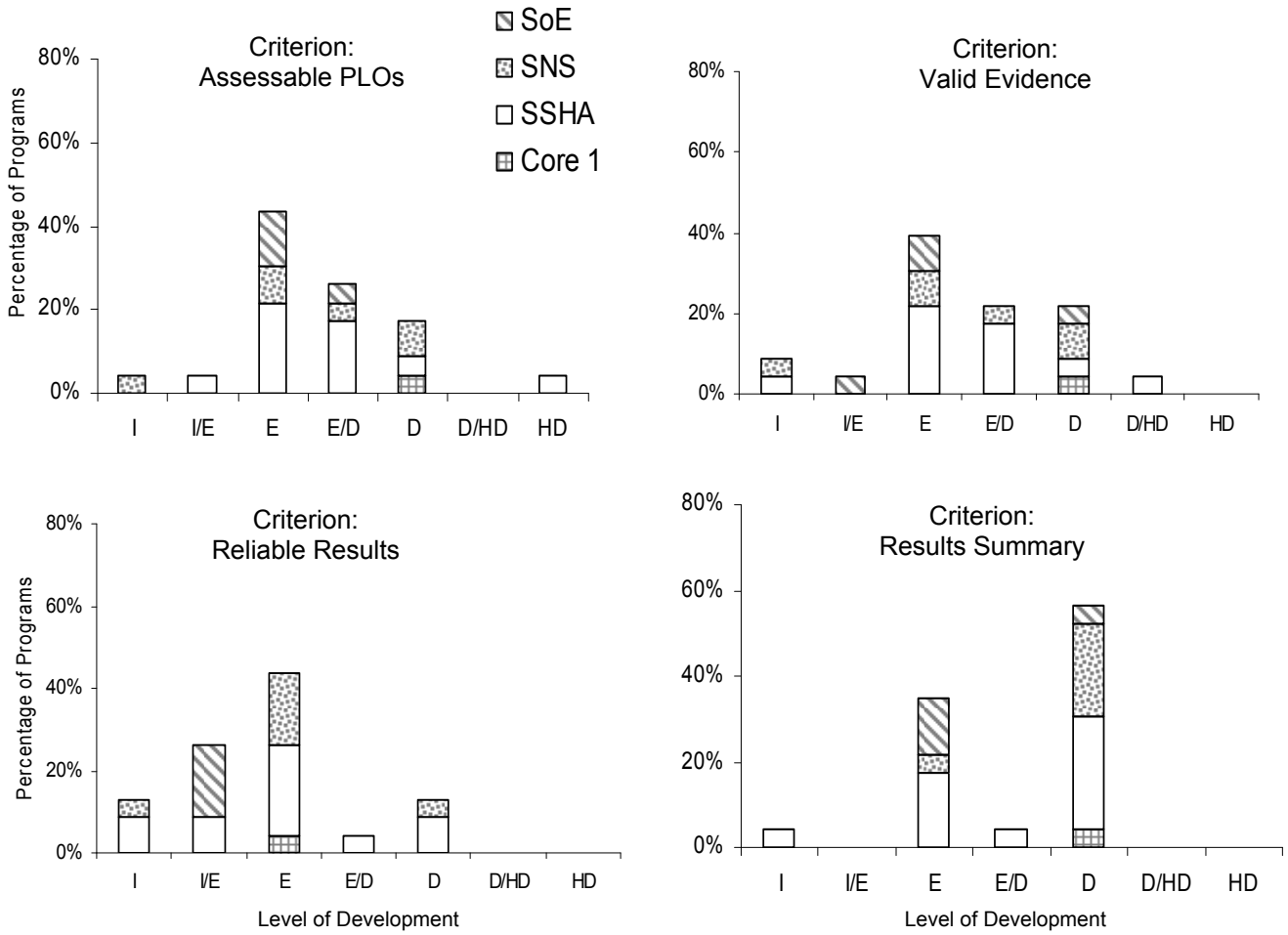
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



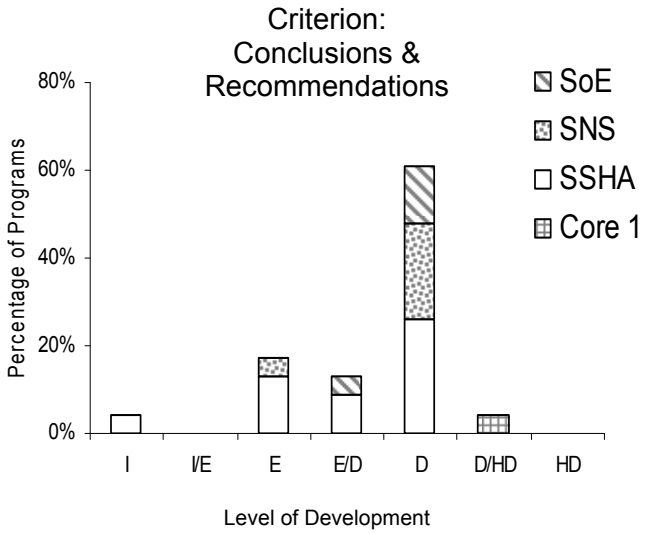


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

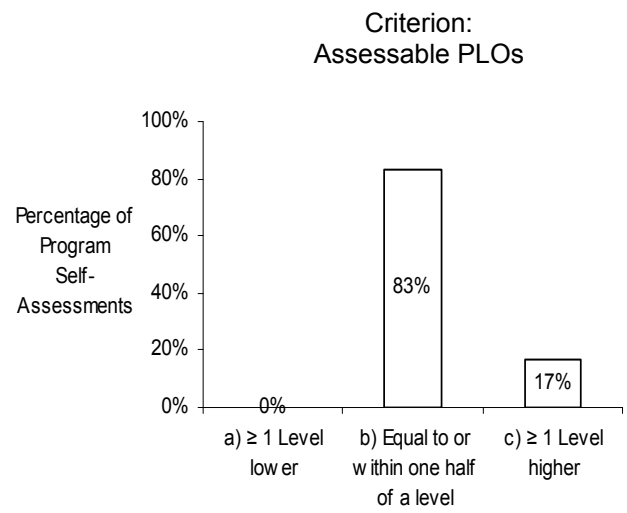
These results indicate that

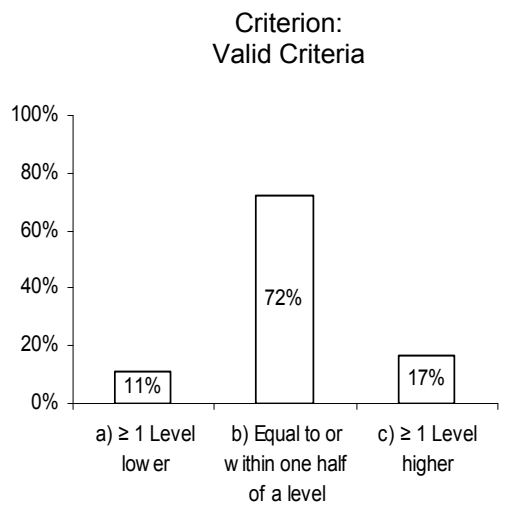
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

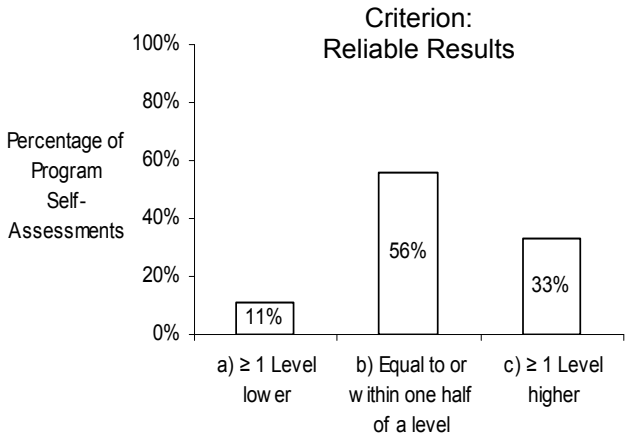
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

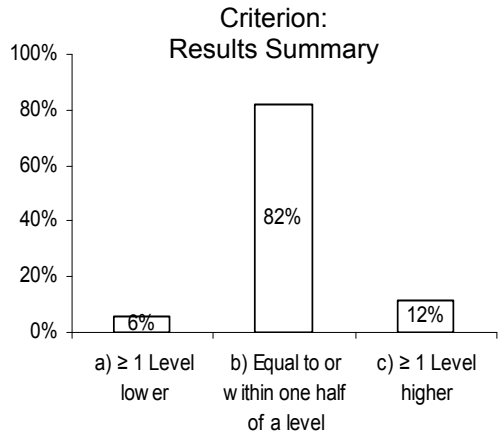
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









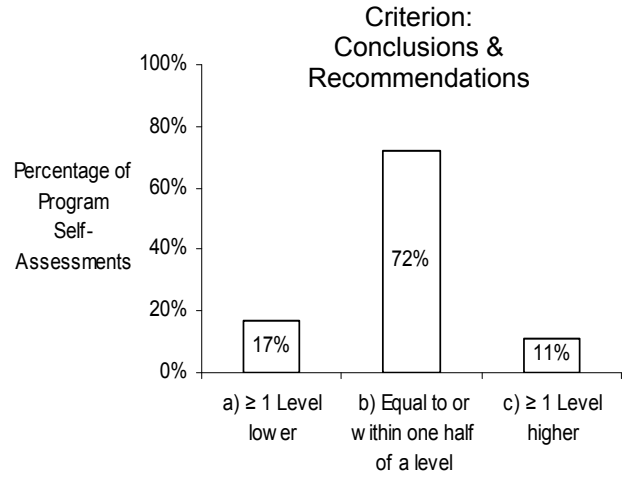


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Tom:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **CORE 1**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**D**	**D**	**E**	**D**	**D/HD**
Primary	D	D	E	D	HD
Secondary	D The rubric is extremely detailed without being impossibly cumbersome.	E/D The PLO assessment does not include indirect evidence.	E As was mentioned in the report itself, the addition of an inter-rater reliability assessment would be a strength.	D Nice balance of tabular presentation and numerical summary.	D There is insufficient consideration of reliability to warrant a rating of HD.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: That this assessment picks up on practices already in use makes it richer, more detailed, and more substantial than most.

Secondary Reviewer: The scoring rubric is exemplary; there is careful attention to exactly the right details.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) As already identified, there is not a satisfactory degree of reliability between the different raters. This may be a consequence of using the rubric to establish sub-scores rather than using a rubric to guide holistic scoring. See below.

Secondary Reviewer: (2) CORE 1 should work to establish an indirect line of evidence that provides information complementary to the direct evidence. (3) In addition, a more formal treatment of rater reliability would be desirable.

6) Any outstanding thoughts or questions?

Primary Reviewer: Perhaps the instructors' expectations of student performance on such a complex task as the cumulative essay are too high. Danger of teaching "to the test:" "[T]o develop means of better preparing students for writing the Cumulative Essay." Shouldn't the task be to help students better integrate disciplinary approaches? I.e., focus less on the outcome and more on the real purpose of the assignment. Similarly, there's a danger of being overly concerned about "consumer preferences": "such that in the Cumulative Essays we have some implicit yet reliable evaluation of what course foci were of particular interest, and therefore use." Maybe just says what needs better presentation, more foundational work, betters scaffolding, etc. In other words, student interest is only one component of pedagogical choice; our task is less to respond to students' current interests than to broaden them! On the other hand, the discussion on p. 12 shows engagement in the course purpose, using the success of student work to show how well some ideas are being developed. This is, I think, a much better use of assessment review.

While I think the speculations on what's working and what's not are important, and ones I would have made early in my career, my longer-term experience suggests that some of these are probably unrealistic: "This may be because the essays are rich in ideas but not cohesive or polished enough. Again, teaching the Cumulative Essay piecemeal, and with regard to other, similar assignments, throughout term, may help in encouraging more developed essays later in term. In short, we need to match the execution of the essay with the ideational energy that fuels it, so as to maximize both." Chances are that no scaffolding will yield satisfactory results for most students on the first try. Successfully managing the rhetoric of longer, synthetic papers takes practice. Properly scaffolding the precepts, and giving examples will help, but mastery requires multiple repetitions.

I'm also surprised at the use of numerical rubrics rather than holistic grading. Holisitic grading is the norm in evaluating writing in part because it prevents the kind of micro focus on parts that may be weak by the rubric but that are not, according to the particular plans of the particular essays, as important in some cases as in others. Holistic grading allows a reader to hold in mind the general components, as per the rubric, but judge overall effectiveness according to authorial purpose by a single-digit score.

Secondary Reviewer:

This may be one of those cases where teaching to the test would not necessarily be a bad thing. Consider just one of the criteria, for example, Synthesis. Teaching to the test would include teaching students to include a thesis, include specific claims, develop their ideas clearly, and bring in appropriate supporting evidence. These are not bad things to teach.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

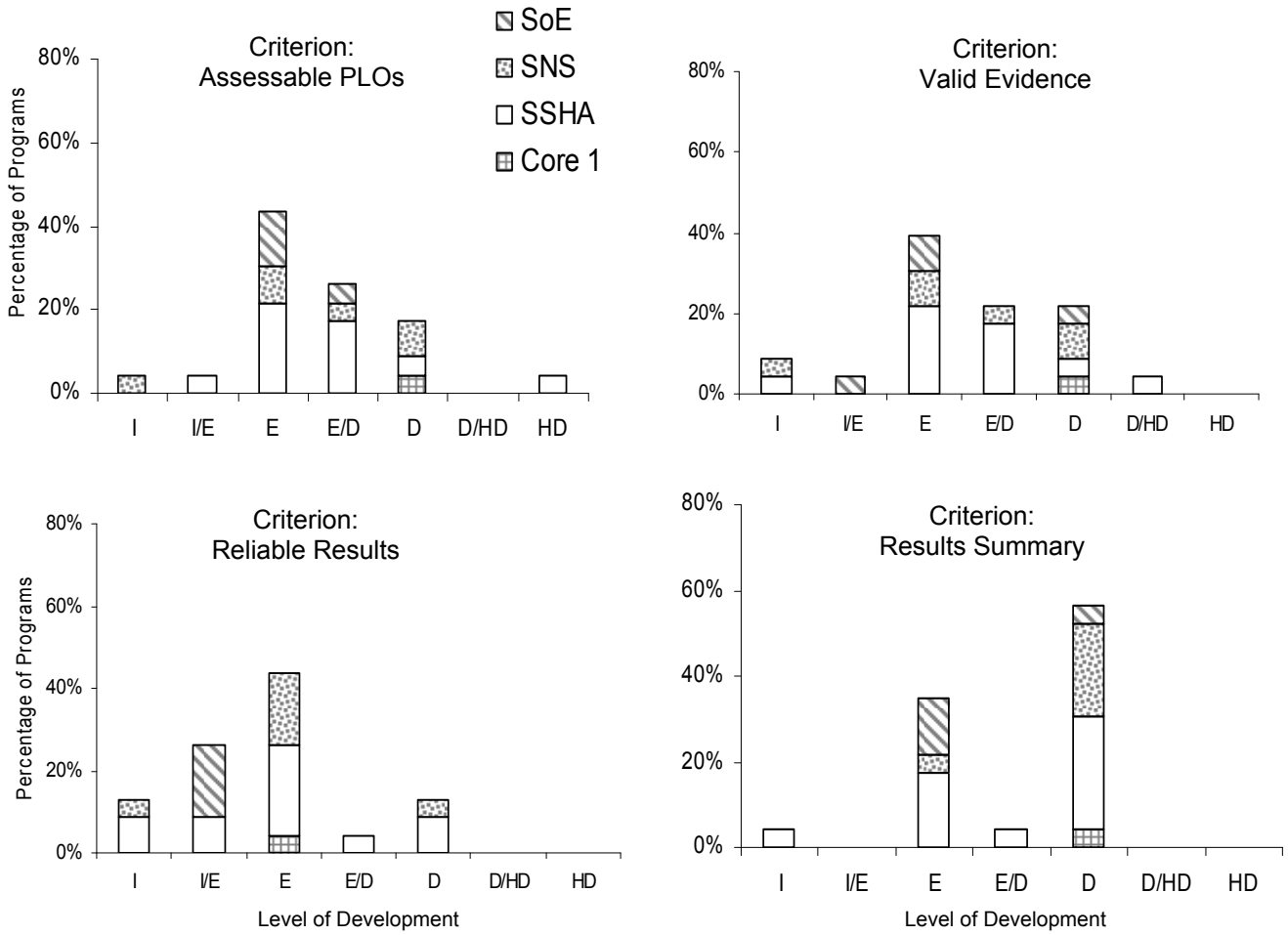
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



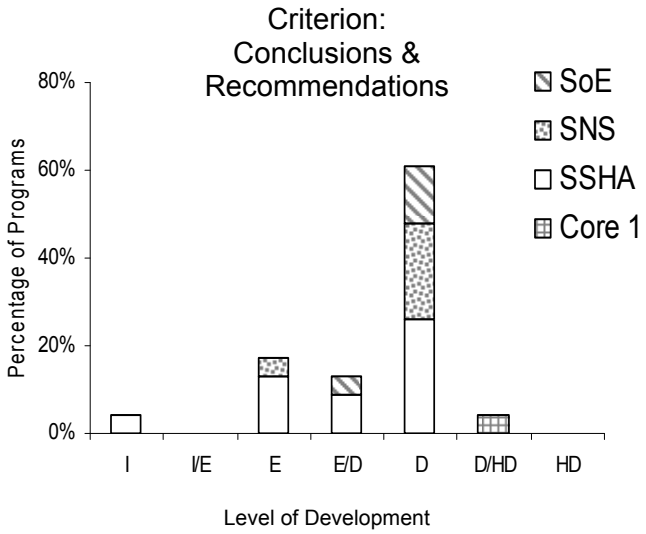


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

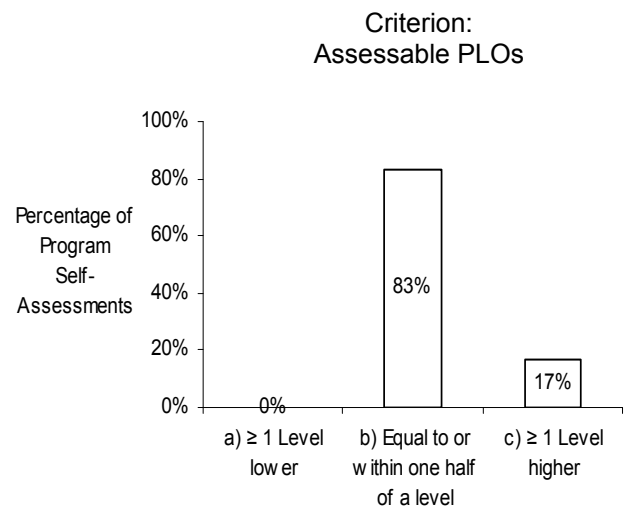
These results indicate that

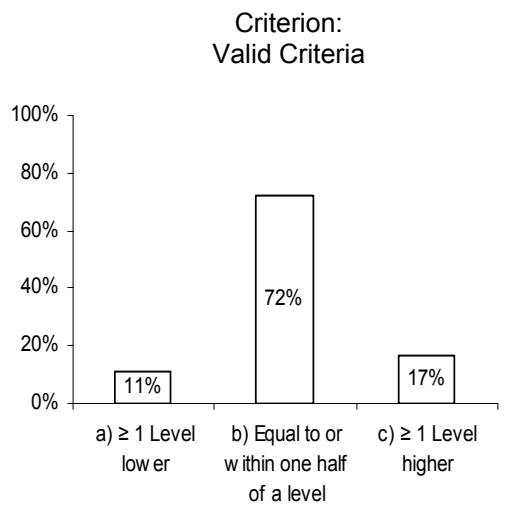
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

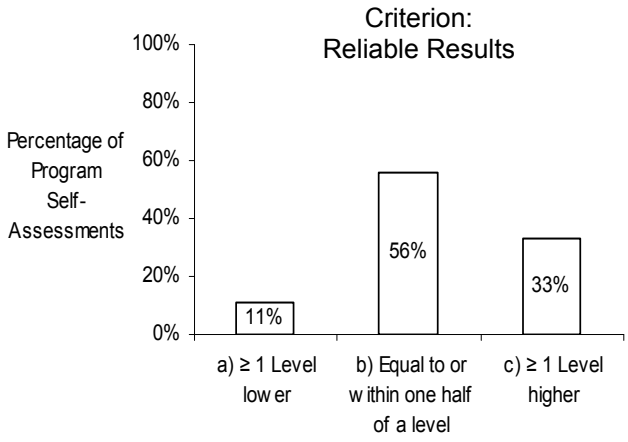
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

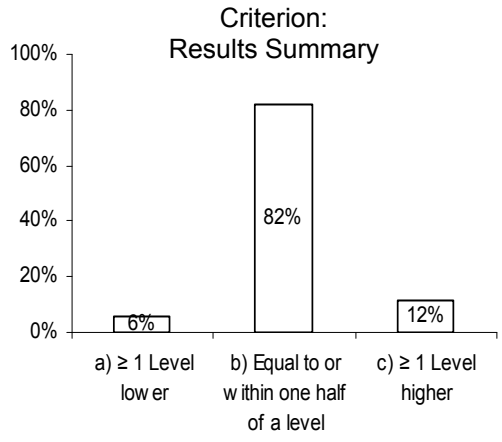
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









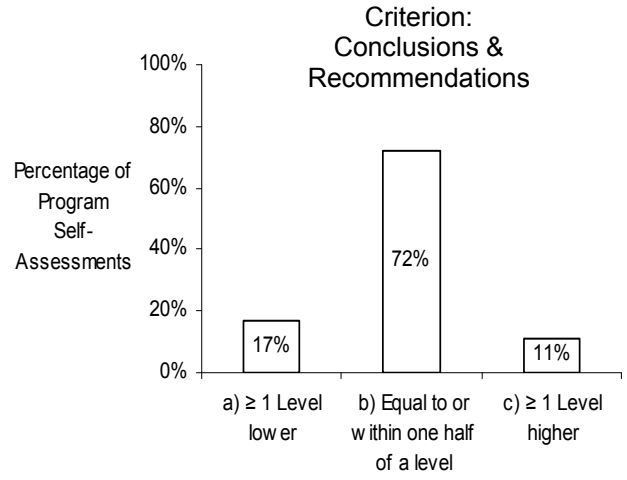


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Ming-Hsuan:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **COMPUTER SCIENCE AND ENGINEERING**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**I/E**	**I/E**	**D**	**E/D**
Primary	E –the action verb is general and assessment criteria are in development; faculty have developed relevant domains of knowledge to focus the evaluation of the PLO	I – faculty are assessing performance in their own classes, presumably based on their own grading, so student achievement is assessed on the basis of instructor-specific criteria	I – there is no calibration of faculty criteria, except through a single individual	D – narrative summaries describing student success as judged by scores are provided and in some cases compared to past versions of a course as a sort of benchmark	E/D – clear recommendations emerge and it seems that the results are likely to be implemented but conflicting statements are made in the report. However, no assessment of the assessment process is provided.
Secondary	E/D – the action verb is general; criteria have been articulated, but standards need further development	E- program criteria are used in assessment; indirect evidence is not included	I/E – although it is not clear if there was calibration of faculty criteria, the FAO reviewed the student work and individual faculty assessment to ensure inter-rater reliability	D- Narrative summaries include percentages of students meeting specific goals, but benchmarks need further development	E/D – Clear recommendations are made as a result of discussion by the CSE faculty as a group, although it is not clear how changes related to preparation of students in non-CSE courses will be addressed; recommendations for improvement of assessment are not included in the report

2) A clear strength of the program's assessment efforts.

Primary Reviewer: One clear strength is the involvement of a large fraction of the faculty (all?) in the assessment process, providing opportunity for productive discussions regarding student performance and further development of a programmatic perspective on what students should know and be able to do at graduation as a result of the program's curriculum.

Secondary Reviewer: : The Results Summary is well developed, with substantive reports provided by each instructor, with detailed descriptions of evidence collected and student performance (including quantitative information).

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) As a faculty, work to develop a program perspective on student learning with respect to this PLO that is grounded in the skills and knowledge students will be able to demonstrate at graduation as a result of their educational experiences across the program. This could naturally evolve from the comprehensive reviews of student performance reported here. A rubric is a very useful tool for formalizing these criteria (the attributes of student work) and associated standards (levels of performance). ABET provides resources for developing rubrics, see - http://www.abet.org/assessment.shtml#Assessment%20rubrics.

(2) The CSE faculty also need to build indirect evidence of student learning into their assessment process. One potential source of evidence is SoE's exit survey for graduating seniors.

Secondary Reviewer: The criterion needing the most development is "reliable results".

(1) It would be helpful for the faculty to meet as whole to develop an assessment rubric with specific criteria and standards laid out. (Specific criteria have been identified. It is the rubric and clear set of standards that appears to need development.)

(2) As part of development of the assessment rubric, it might be helpful for the faculty to review materials collected for the current assessment exercise

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

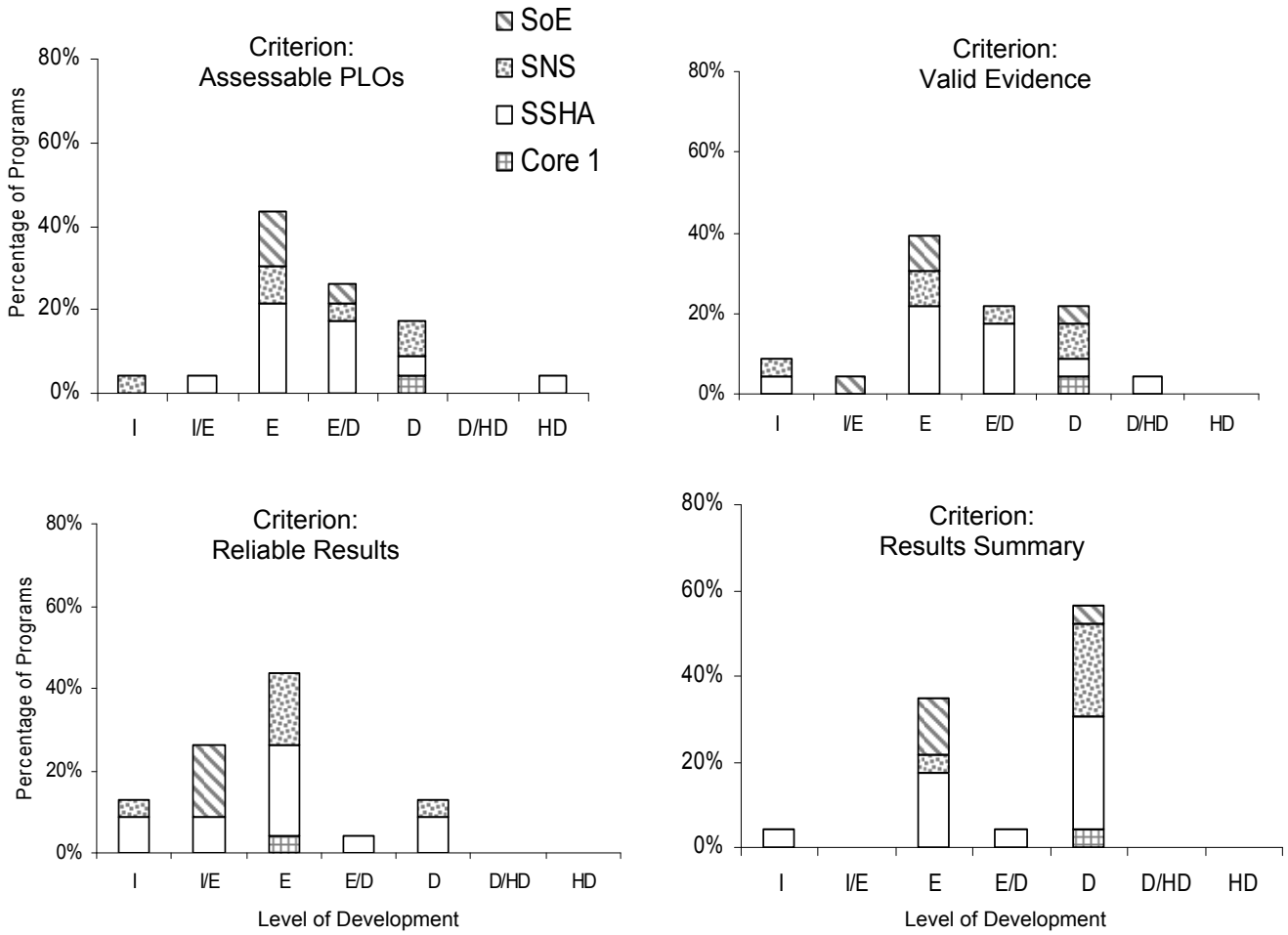
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



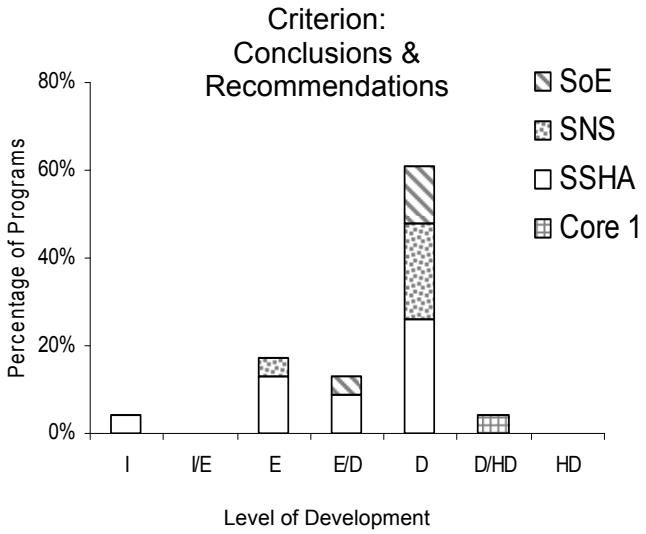


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

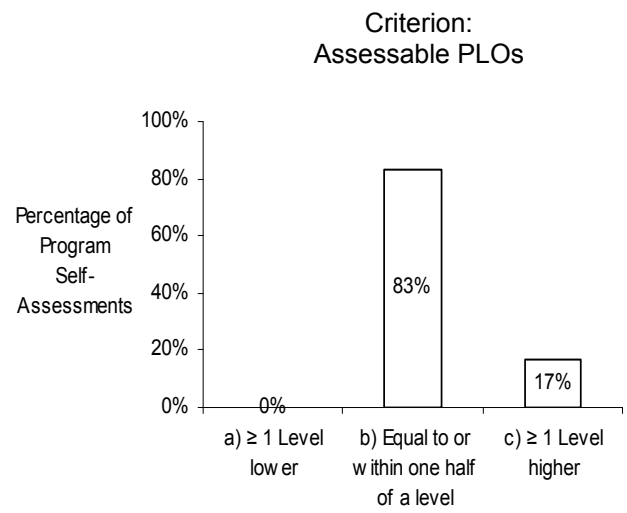
These results indicate that

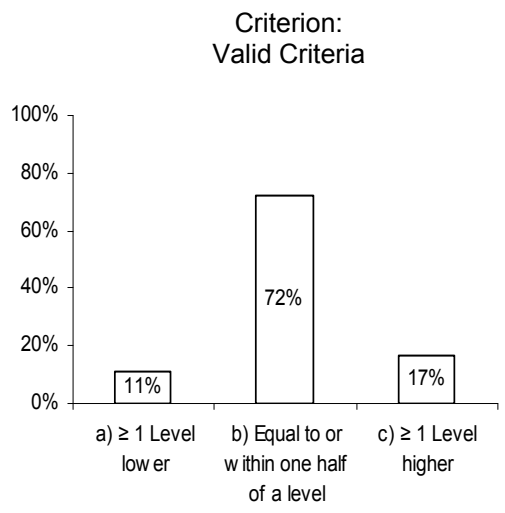
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

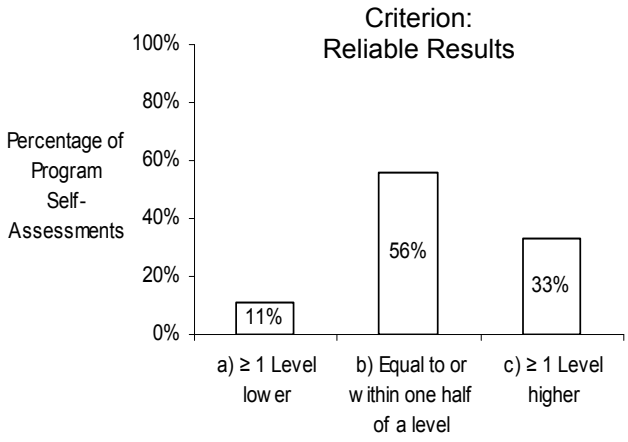
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

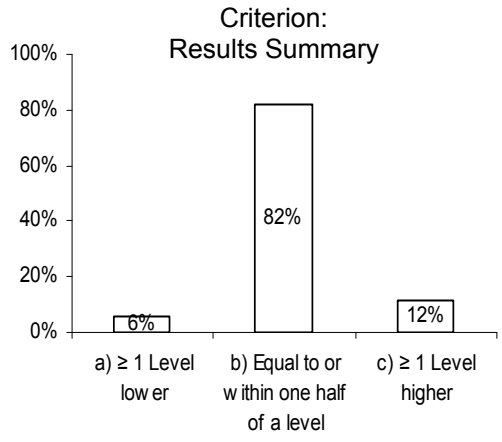
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









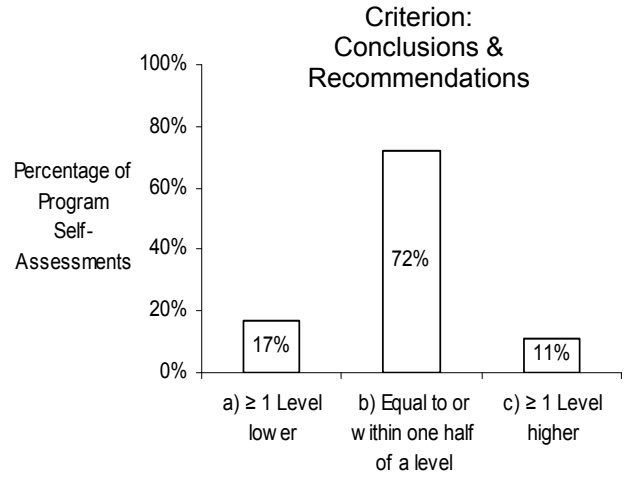


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Tom:

Please find attached feedback regarding Environmental Engineering's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **ENVIRONMENTAL ENGINEERING**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E/D**	**E**	**I/E**	**E**	**D**
Primary	E/D – the action verb is general; specific criteria have been articulated, but standards need further development.	E-General agreement has been reached by faculty on types of evidence to collect, however the evidence needs to be further aligned with the PLO. Indirect evidence is not included.	I/E – Reviewers do not appear to have been calibrated to apply assessment criteria in a uniform way, however the FAO reviewed the evidence to ensure inter-rater reliability.	E-Results are included in the form of a rubric mapping the relevance of each course to the PLO, and the performance of students in each class, however a rubric clearly delineating results for each line of evidence is not included.	D – Clear recommendations for improving student learning and assessment, along with discussion of the validity and reliability of the results, are included.
Secondary	E/D – action verb identifies observable, measurable results but assessment criteria need developement	Agree	I/E – personal criteria are used and then validated by the FAO; the criteria and standards used to judge the quality of student performance is not shared across the faculty	E – average student performance for each line of evidence is reported but have no idea of the distribution of performance.	Agree

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A clear strength of the program is mapping of relevance of specific courses to the PLO, as well as the specific conclusions regarding mechanisms for improvement for student learning and assessment.

Secondary Reviewer: A clear strength is the collection of lines of evidence from multiple courses, which supports moving toward a programmatic perspective on student learning that reflects explicit expectations, across the faculty, describing the knowledge and skills students will demonstrably possess at graduation.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer:

1) Further discussion among the faculty regarding valid evidence to collect and calibration of faculty criteria would be useful.
2) It might be helpful to more closely align the lines of evidence with the PLO. Although the relevance of each course to the PLO are nicely mapped, it is not clear that the lines of evidence chosen (homeworks, midterms, quizzes, research report) map directly to the various aspects of the PLO.
3) Along the lines of comment #2, it might be helpful to more closely align CLOs to the PLO. The program has already noted the need to include teamwork in CLOs.

Secondary Reviewer: (4) Additionally, as the program noted it should work to integrate indirect assessment.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

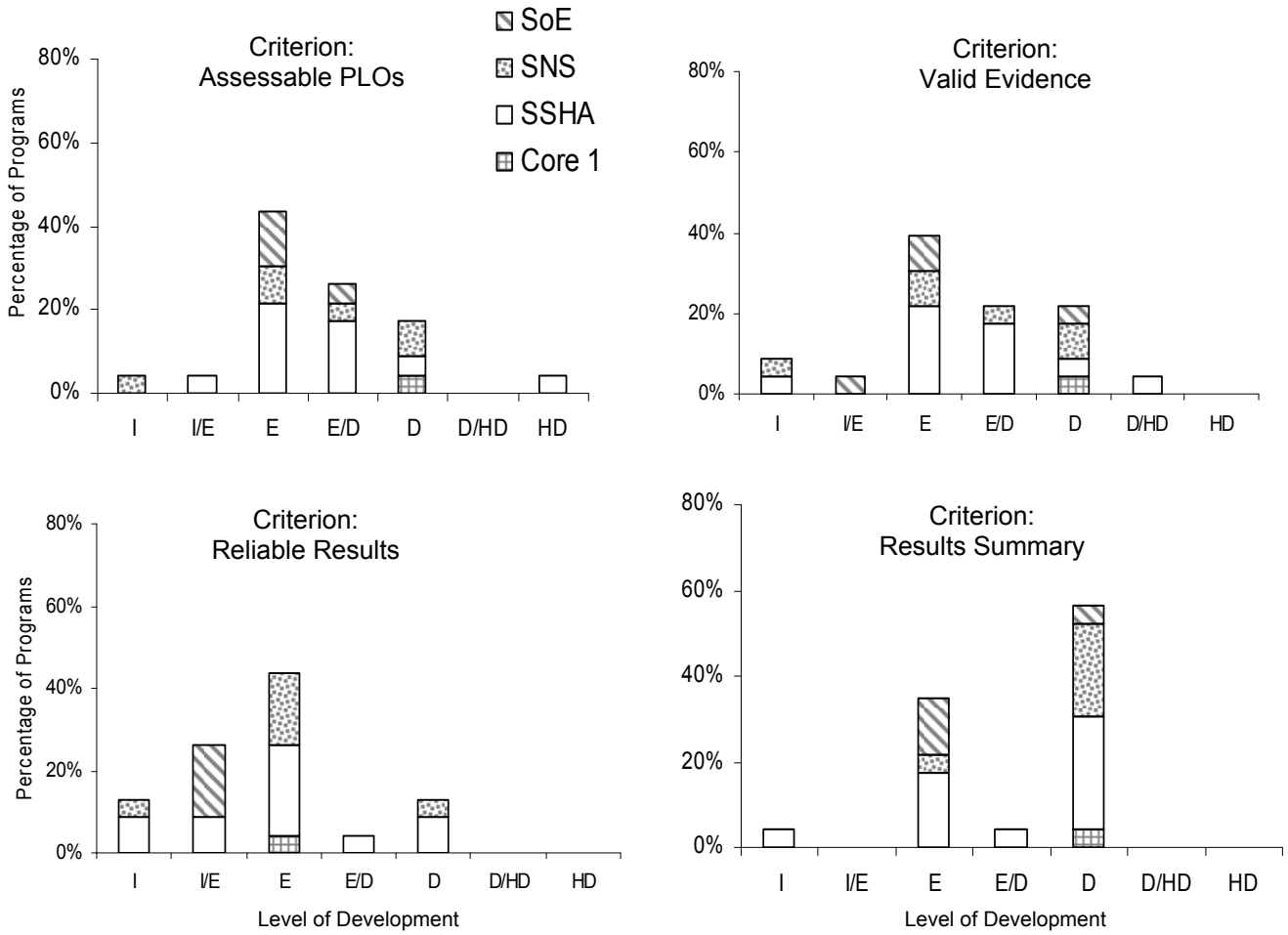
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



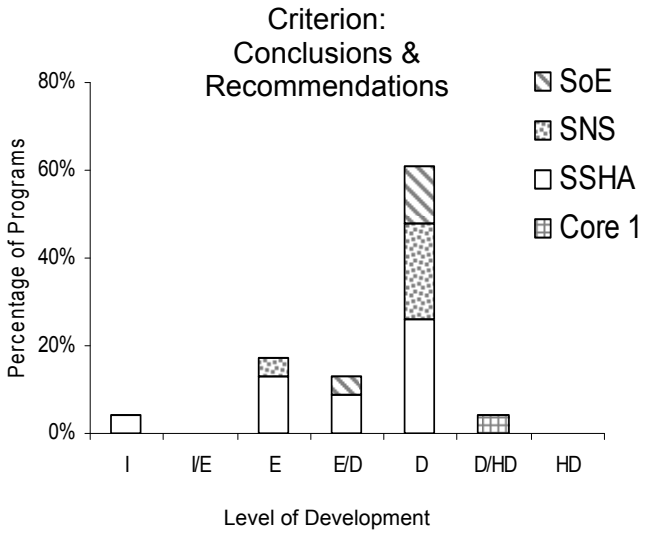


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

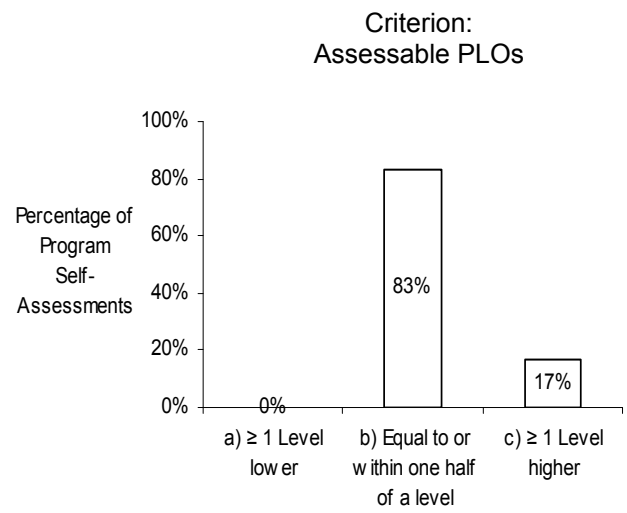
These results indicate that

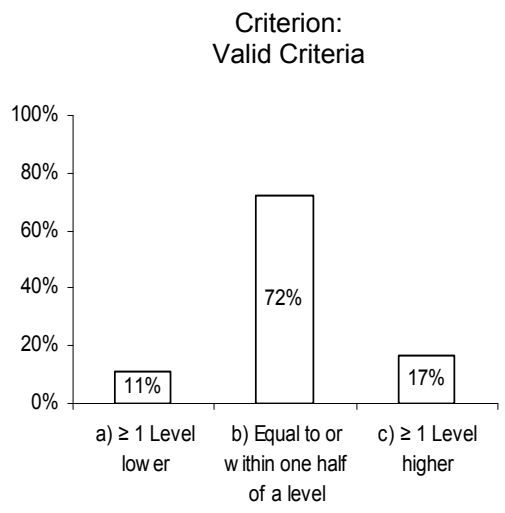
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

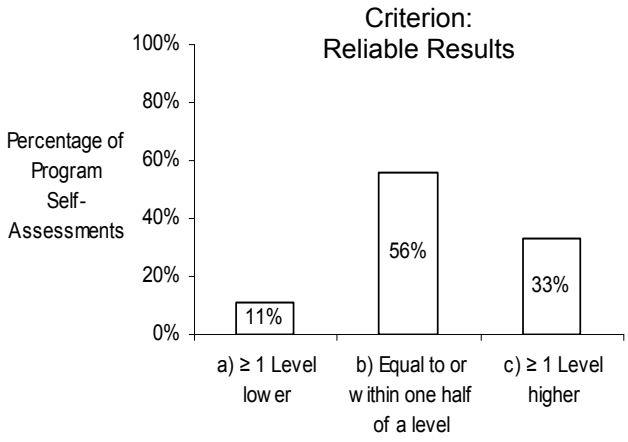
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

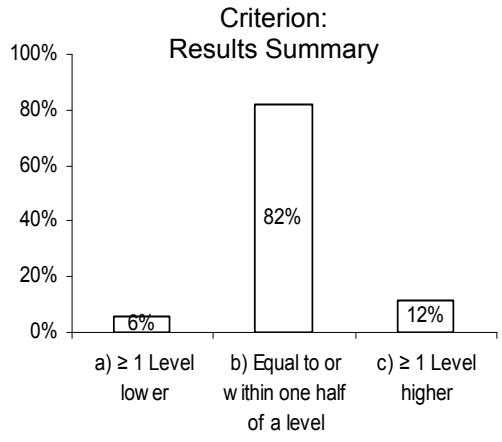
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









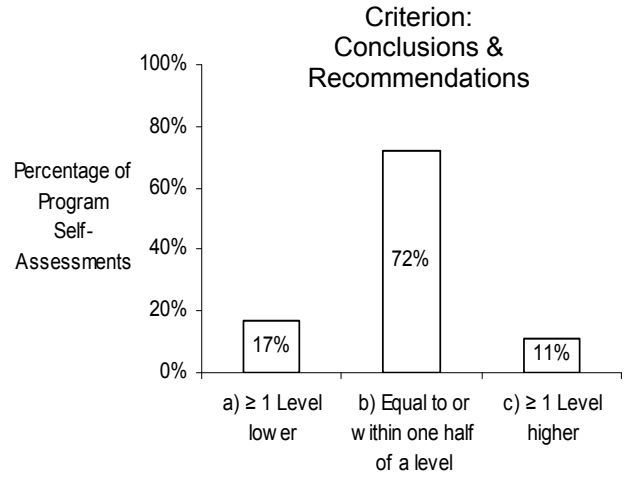


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Tom:

Please find attached feedback regarding Environmental Systems' 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **ENVIRONMENTAL SYSTEMS GRADUATE GROUP**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**E**	**I/E**	**E**	**E/D**
Primary	E- The PLO gives specific criteria indicating how students can demonstrate learning. A rubric has not been developed articulating standards for each criteria.	E- Faculty have general agreement on the types of evidence to collect, but indirect evidence is missing. Evidence needs further alignment with PLO.	I/E – Although it is not clear if there was calibration between faculty regarding assessment of student work, the FAO reviewed the student work to ensure inter-rater reliability.	E-Student performance results are included for the class proposal presentations, but the data could be more complete and more specifically address the PLO.	E/D-Some implementable recommendations are given regarding improving learning and assessment, and the results have been shared with the ES faculty.
Secondary	Agree	I/E – although two faculty reviewed the student work the review criteria are personal rather than programmatic	Agree	Agree	E/D – but closer to E. Conclusions address assessment but not student learning.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The PLO is very specific, with assessable criteria.

Secondary Reviewer: The faculty are instructing and assessing for an essential set of skills early in the program; this should really help the program support student development in these critical areas.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: It would be very helpful to develop a rubric of criteria vs. standards, and calibrate all faculty to the standards. Additional evidence should be provided to fully support assessment of this PLO – at present the evidence does not assess the data analysis and formulation of conclusions aspects of the PLO. The PLO assumes a different level of mastery for PhD and MS students (independent vs. supervised work), but it is not clear how the assessment itself measures the degree of independence of students in completing the work.

Secondary Reviewer: I agree with the primary reviewer, noting that the ultimate goal is to examine student proficiency with respect to the PLO at the time of graduation. In this context, ES 200 establishes an early benchmark in student skill levels with respect to this PLO, providing the program the opportunity to examine development of these skills over the course of their education.

Additionally, as ES noted, it should also work to integrate indirect measures into its assessment. Since the program indicates that rubrics will be provided to students to help guide their work, the indirect measures could in part take the form of student self-assessments against this rubric. This could be particularly useful if the students will continue to meet this rubric throughout their graduate schooling.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

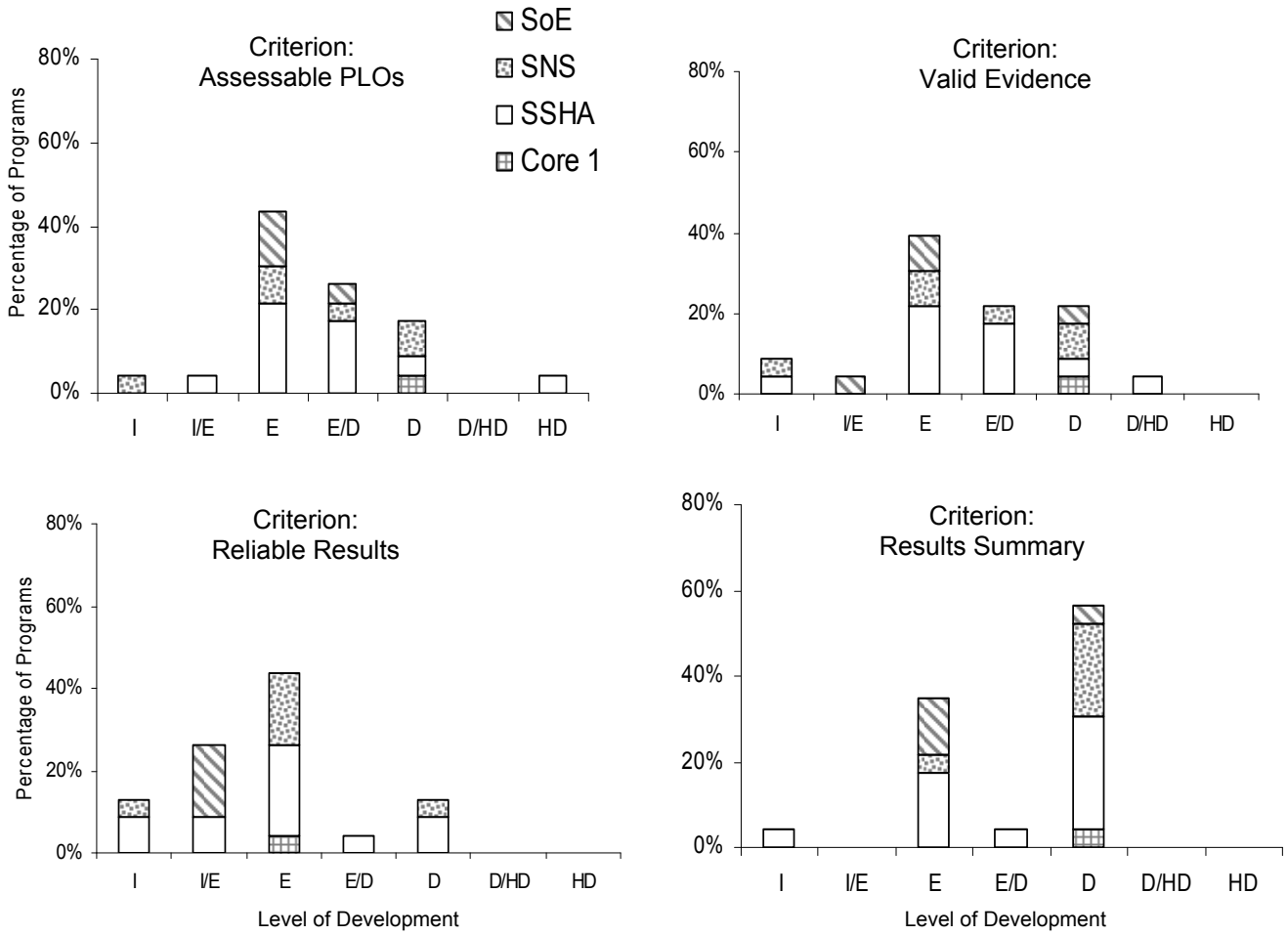
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



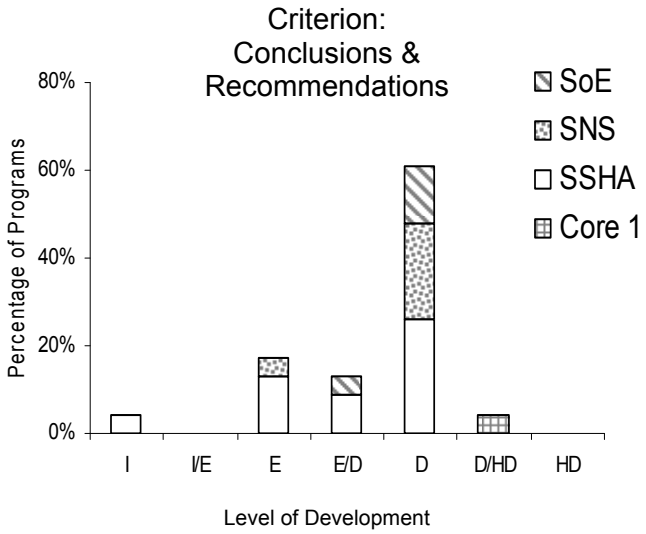


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

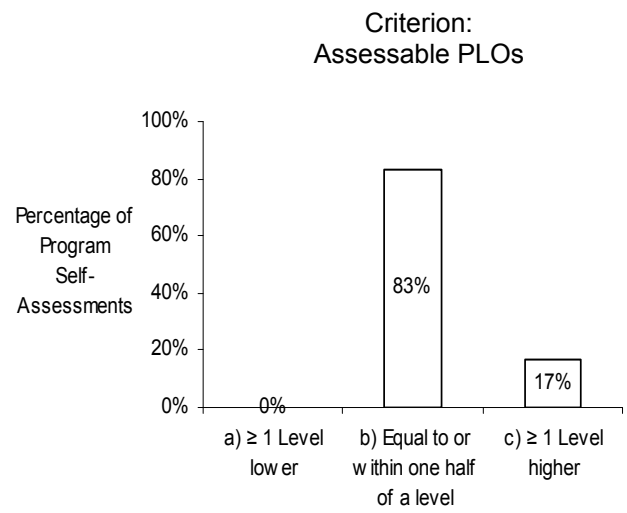
These results indicate that

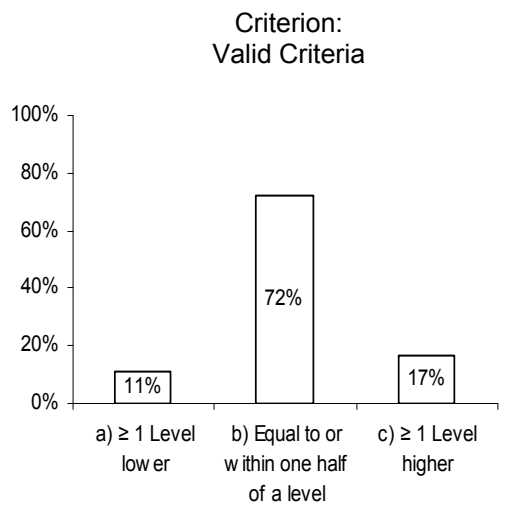
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

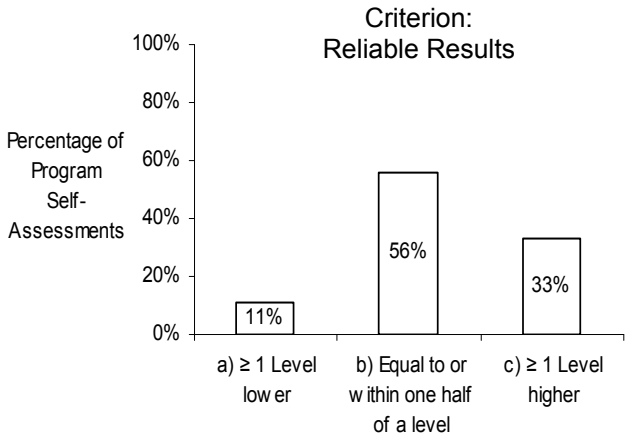
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

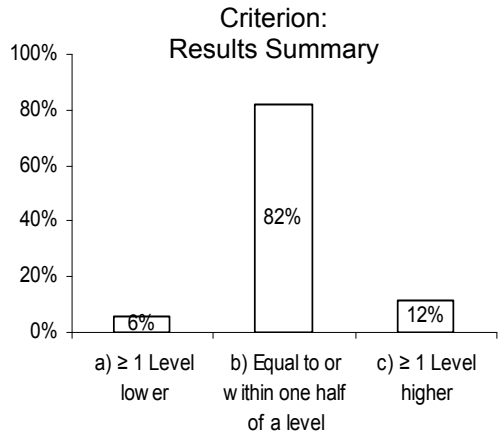
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









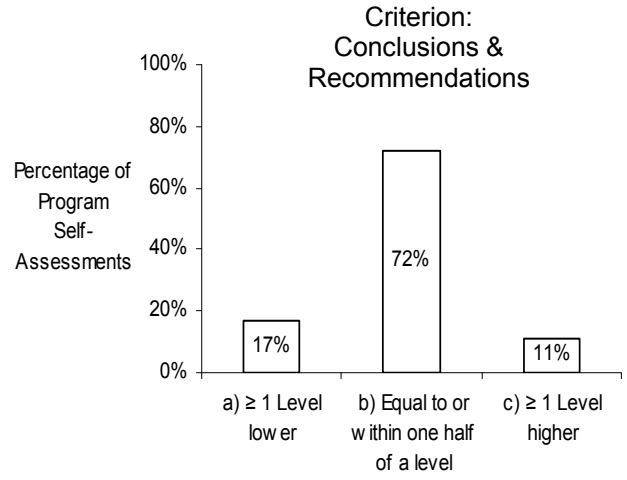


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Lara:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **EARTH SYSTEMS SCIENCE (ESS)**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**I/E**	**I**	**I**	**E**	**E**
Primary	I/E – PLOs have some specific nature. However generally vague. Neither rubric nor clear assessment criteria has developed. Only exit survey was mentioned as a mean of assessment	I – It is not clear that valid evidence for assessing the PLO was collected. Evidence needs to be aligned with PLO.	I – the exit survey results are not really aligned with PLO.	E – the survey results are included, but those don't assess student work or performance relative to PLO.	E – identifies some implications and recommendations, but should discuss how direct and indirect evidence will be collected.
Secondary	Agree	Agree	Agree	Agree	Agree

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A clear strength of the program's assessment efforts is that 57% (13 out 23) ESS majors were able to participate in the program survey.

Secondary Reviewer: Agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) Continue to develop a shared understanding among faculty with respect to the skills and knowledge that support each PLO. For each PLO, articulate this in a rubric that can be used by faculty. Related to (1), it is obviously important to develop meaningful assessment methods to measure student attainment of PLO. Does a

survey provide sufficient evidence for assessing this PLO? Does the assessment method include direct evidence? How might EES best measure a level of “Knowledge of major concepts of ESS”?

Secondary Reviewer: Develop and implement direct assessment methods and at least a single complementary form of indirect evidence.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

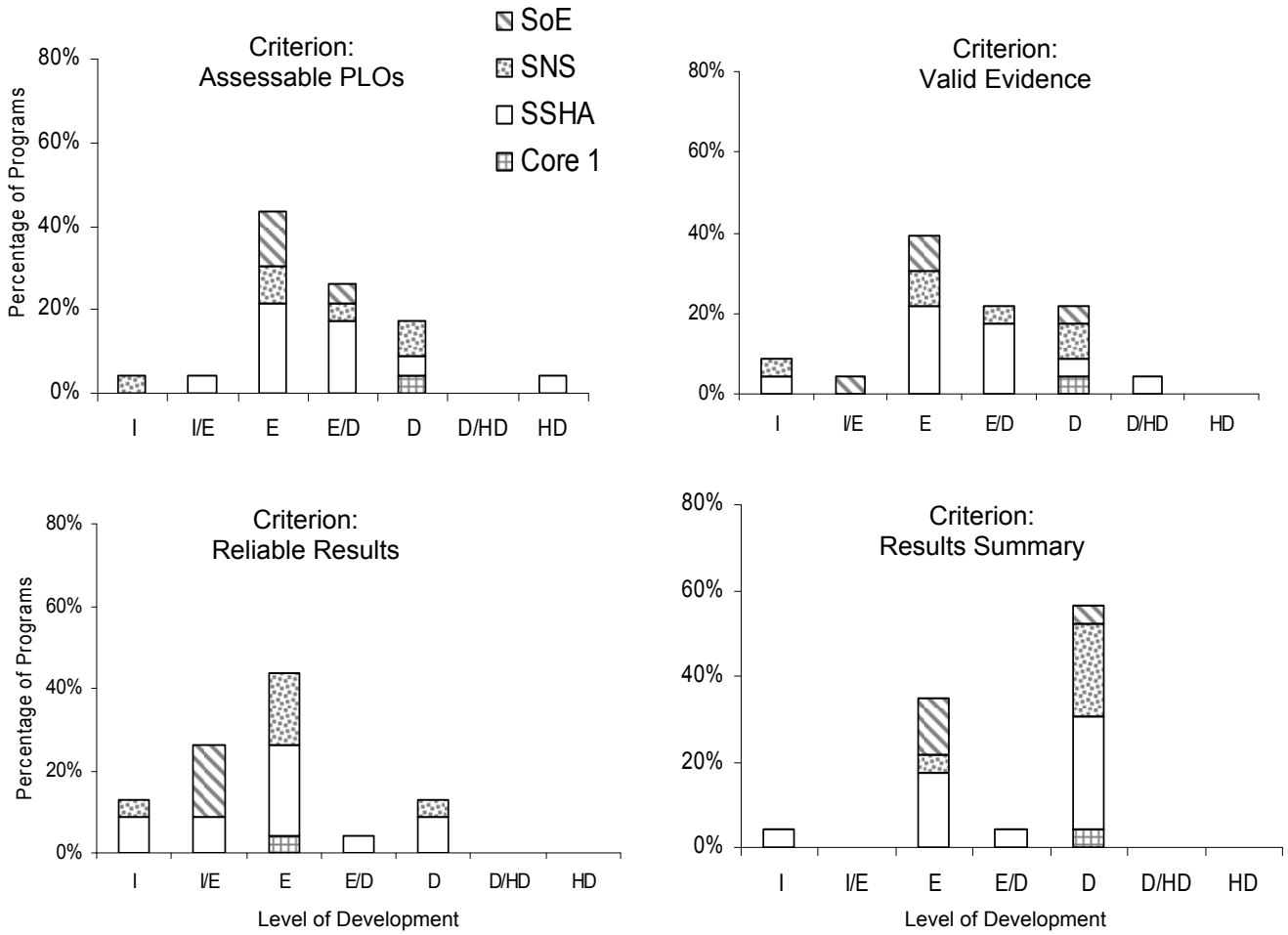
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



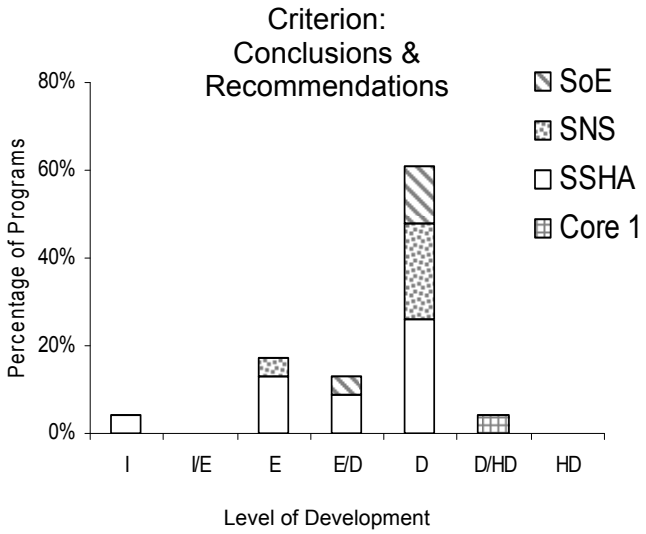


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

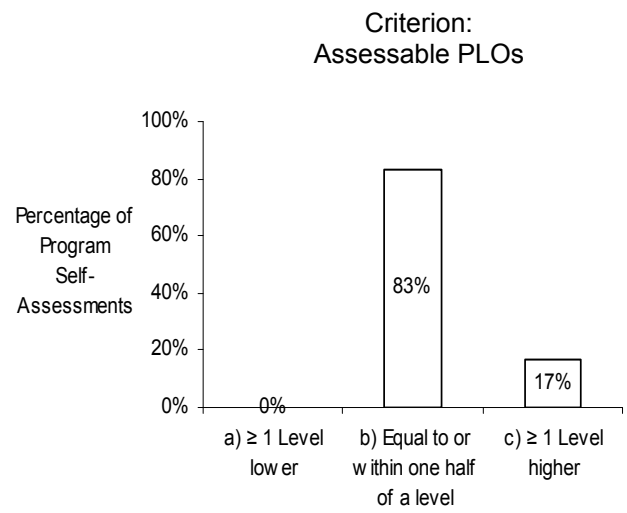
These results indicate that

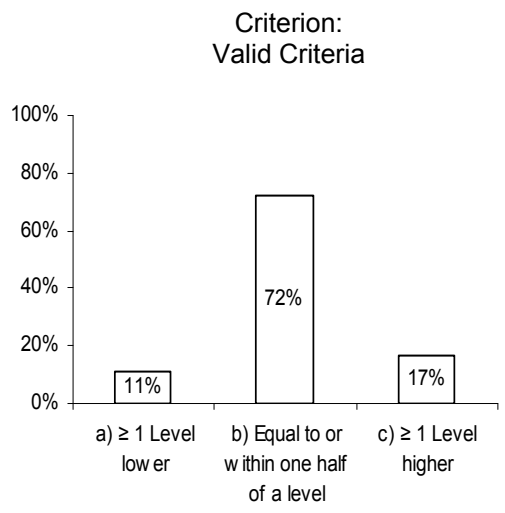
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

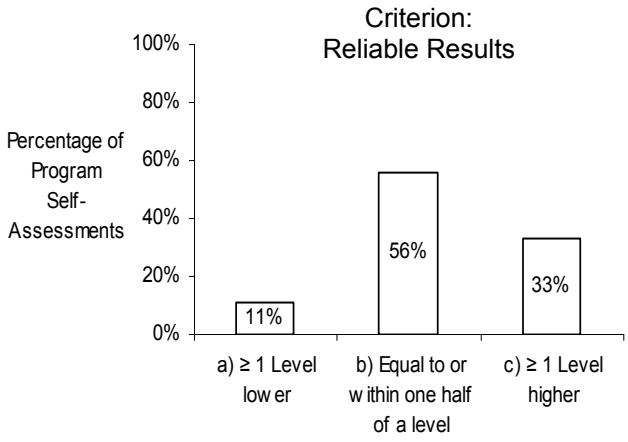
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

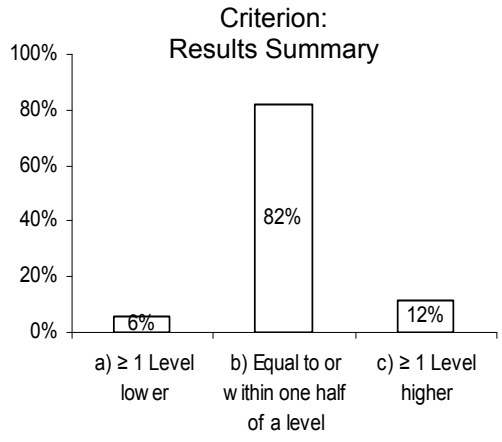
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









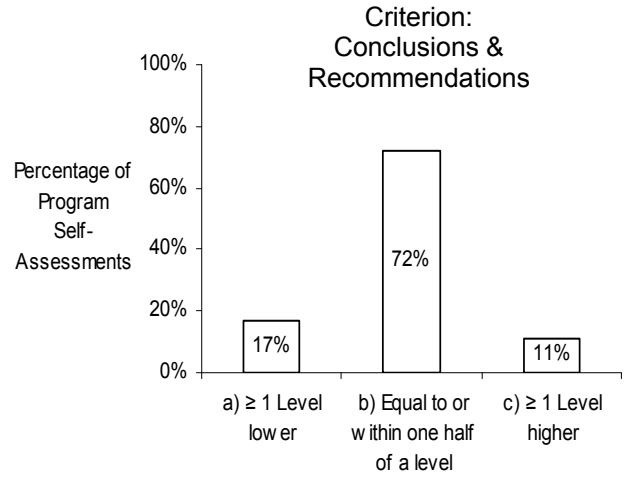


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Sholeh:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **HISTORY**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**D/E**	**E**	**E**	**D**	**D**
Primary	**D** (directive, measurable PLO which might be streamlined for better definition)	**E/D** (relevant and rich holistic evidence, although it can be drawn from a wider variety of students; also, rubric can be developed further)	**E** (reviewers agree on assessment criteria though can be calibrated more actively)	**D** (results delineated in tabular format)	**D** (useful reflection on the exercise and implications for future applications)
Secondary	E (even holistic evaluation needs to articulate clear criteria for accomplishment. These criteria must then be applied in the context of the entire assignment, not parsed by sub-categories)	E	E	D	D

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The program appreciates the intricacies of student written expression as it pertains to the PLO being investigated, is willing to reassess means of measuring the PLO, and will adapt its curriculum to better facilitate student attainment of the PLO.

Secondary Reviewer: Agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: There's need for broader sampling and categorization of student work, for better justification of why History 191 is the best place to conduct such assessment, and for a more developed rubric (with respect to what the discipline expects in critical narrative).

Secondary Reviewer: I sympathize with the reviewers concerns about applying a rubric to narrative; if a rubric prevents holistic evaluation of complex essays, then it's not doing its job. That said, the analogy to professional ms review is instructive. While reviewers are asked to give an overall evaluation of a specific ms, they are usually guided by specific questions. These questions are based on the editorial board's expectations. And these questions usually deal with several independent categories (much like rubrics). Good reviewers know better than to handcuff their reviews to these questions, but also keep them very much in mind as they evaluate.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

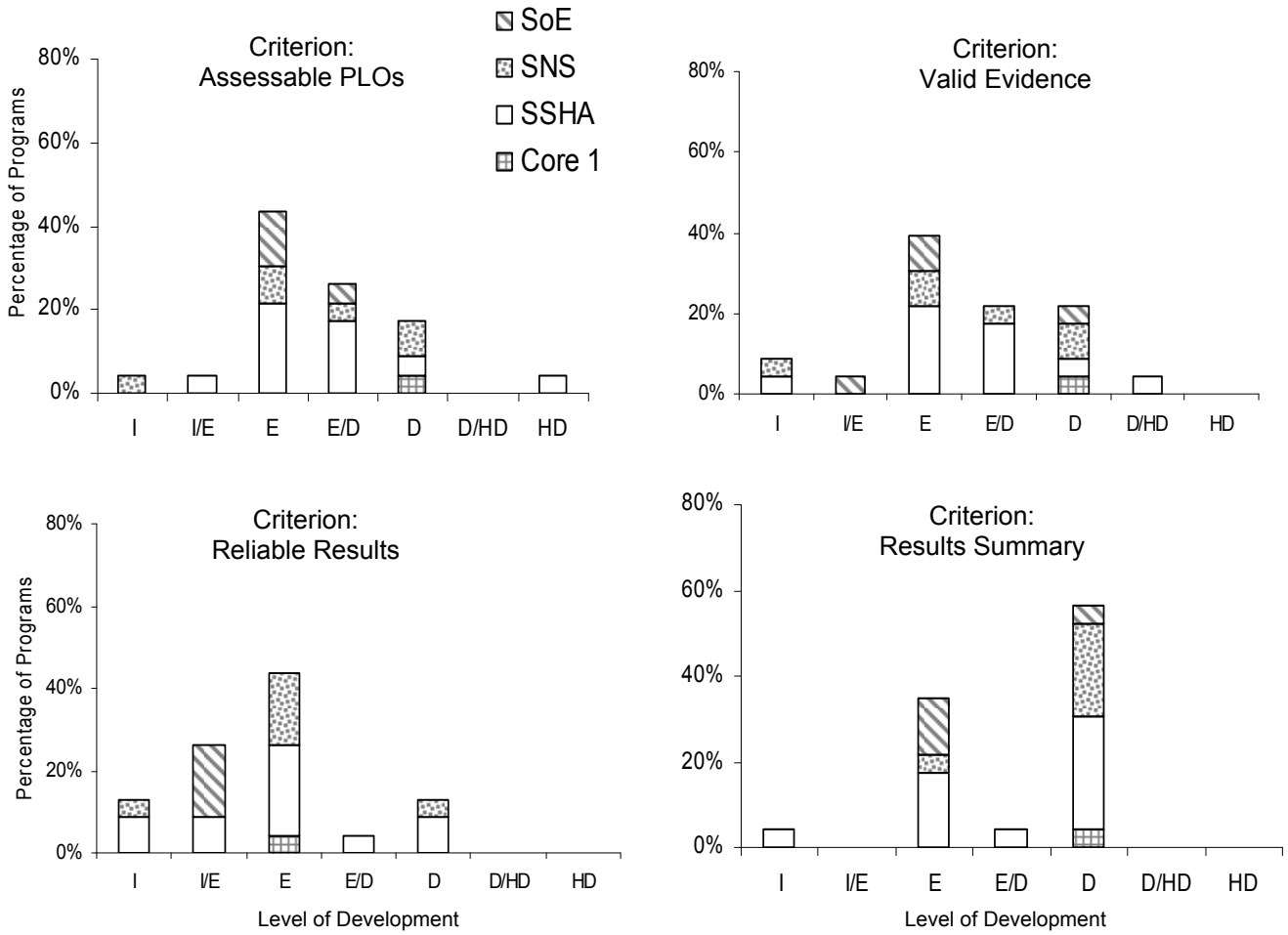
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



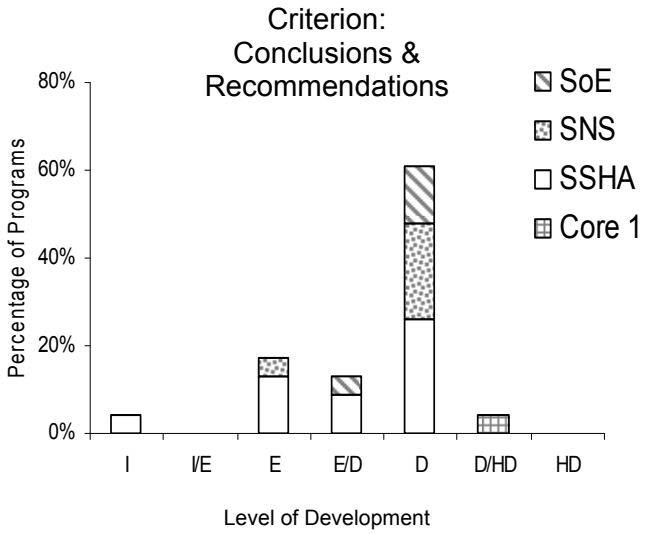


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

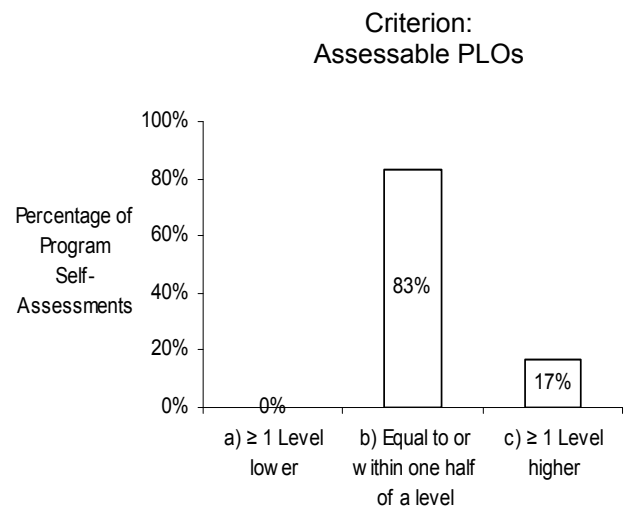
These results indicate that

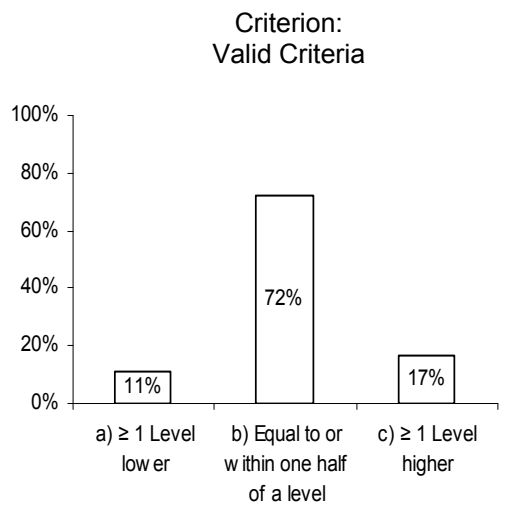
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

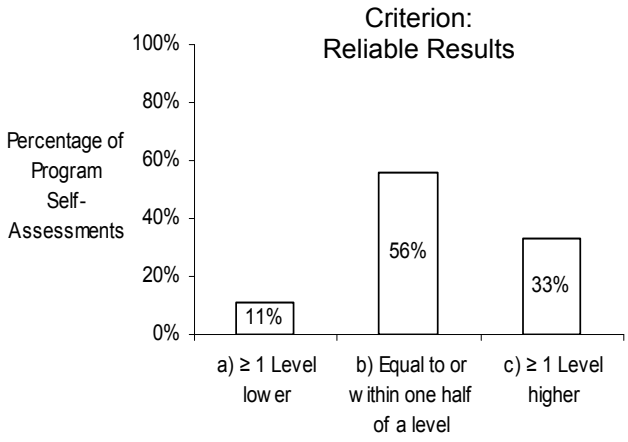
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

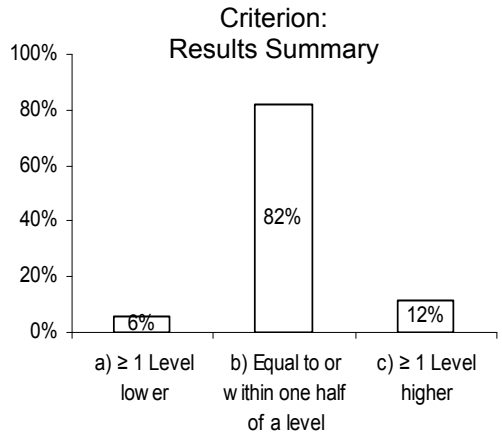
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









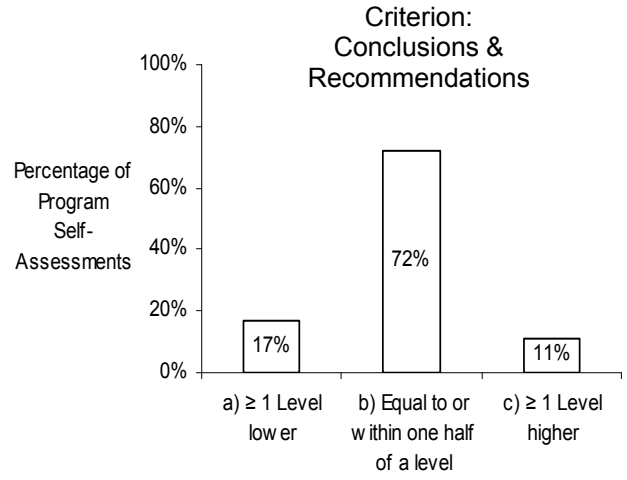


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Jan (and Gregg):

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **LITERATURE AND CULTURES**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**D**	**E/D**	**I**	**I**	**D**
Primary	**E/D** (active, directive PLO, though potentially confusing overreliance on the term (or root) "text" (as in "textual" and "contextual")	**E/D** (rich evidence, although better definition of and rationale for sampling could help define its purpose; also, an elaborate and useful but likely overcomplicated rubric [which would take time to master and apply])	**I** (little explanation of actual process of assessing student work, particularly with respect to calibration or inter-rater reliability, although rubric is instructive)	**I** (no results included)	**D** (lots of reflection on how assessment should evolve, but needs synthesis and direction for specific future assessment efforts)
Secondary	D. The rubric is extremely detailed; although I find some of the terms and ways of thinking difficult, I recognize that they are normal for literary critics.	E/D. The complicated nature of the rubric is probably necessary in order adequately to define the learning outcome.	I. With such a complex rubric, some assessment of inter-rater reliability will be essential.	I.	D.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The program is actively examining the richness and complexity of its curriculum and working toward ways to quantify as well as qualify it.

Secondary Reviewer: The rubric shows careful thought and is grounded in relevant theory of learning.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: As the report notes, there's need for consensus building about how to measure the PLO and perform assessment. The PLO itself could be clarified, so as not to rely on an overdetermined sense of the term "text" (with respect to "textual" and "contextual"); this way, too, it could avoid potential redundancy with another PLO, "judge the value(s) of texts and contexts." The sampling and categorization of student work could be better defined, and student work itself should be included. There could be more discussion of the actual process of assessing student work, the results of that process, and the means of cultivating inter-rater reliability. Overall, the report can seem more concerned with its limitations (what it isn't) than its immediate implications (what it is or could be, specifically).

Secondary Reviewer: The addition of a mechanism for inter-rater reliability is going to be crucial to the success of this assessment protocol. I am somewhat disturbed by the assertion that measuring improvement of individual students is the only valid way to assess student learning. Surely it is also important to evaluate the program in absolute terms: are students who complete it competent in the skills defined by the PLOs in an absolute sense? Indeed, that appears to be the question that the actual PLO assessment addressed.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

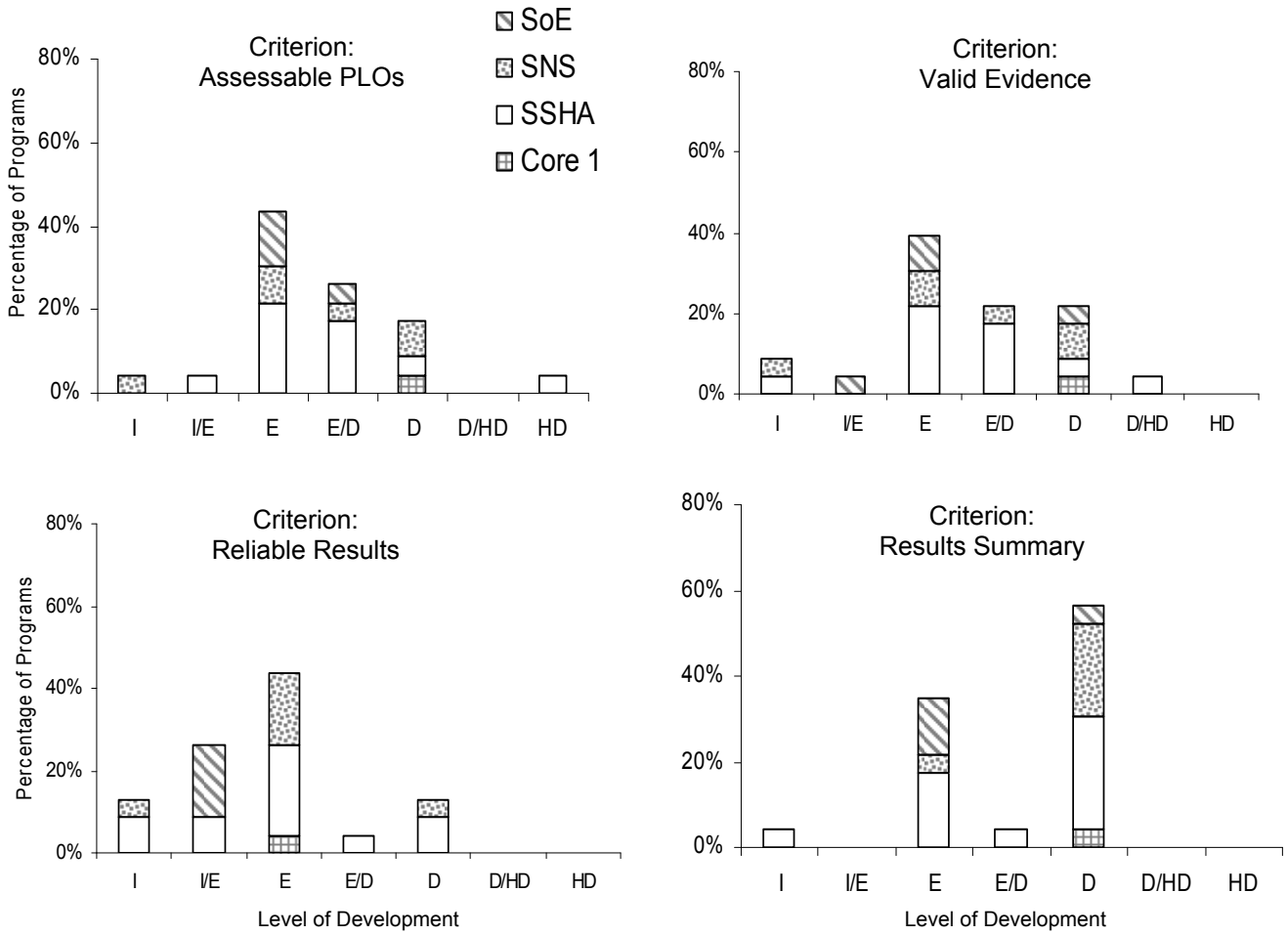
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



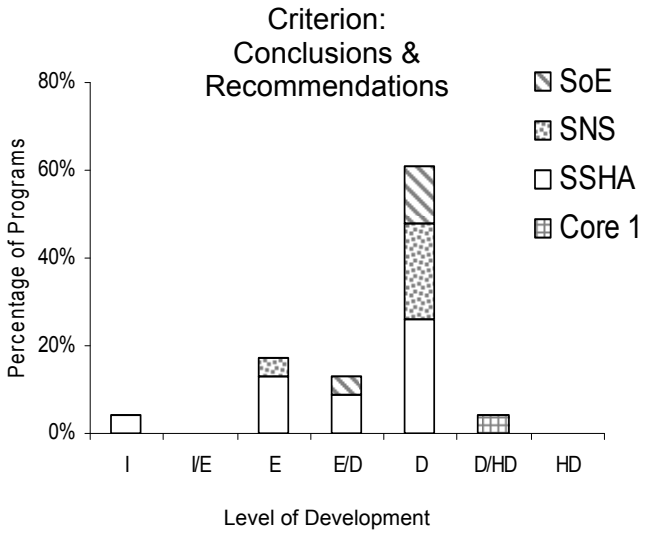


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

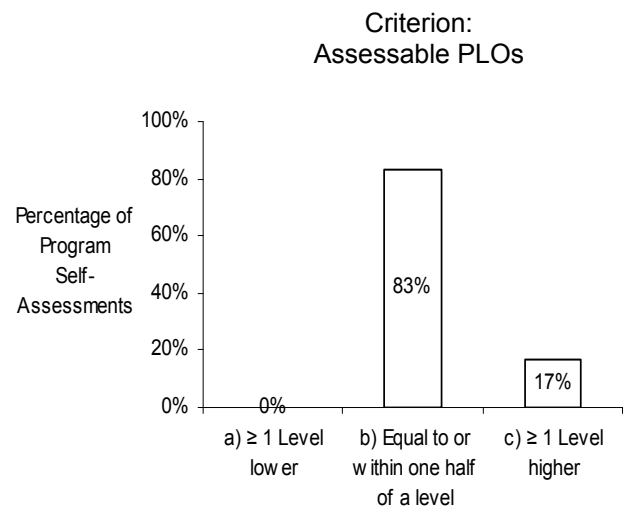
These results indicate that

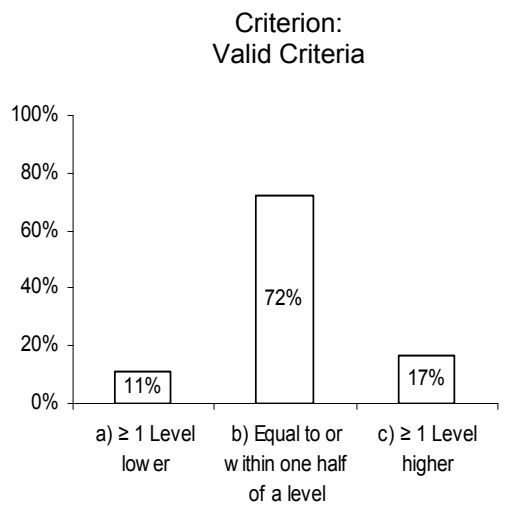
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

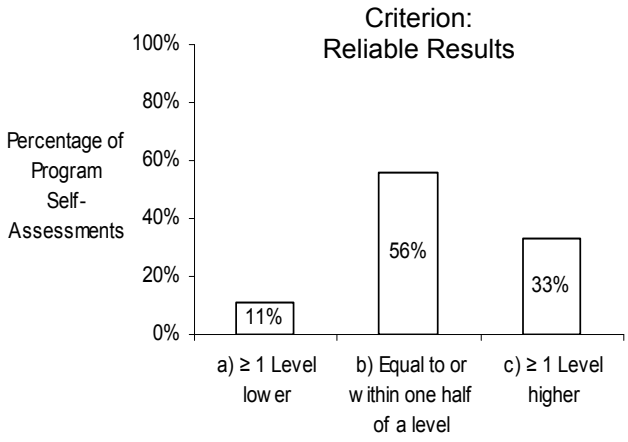
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

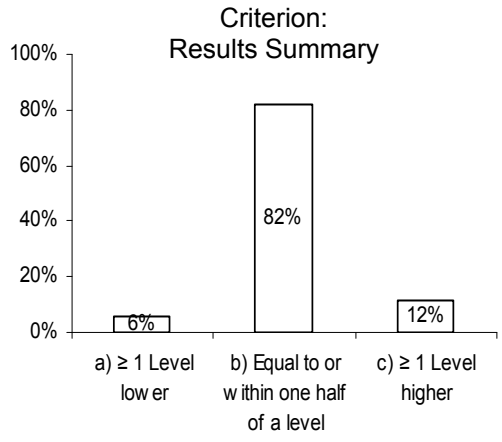
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









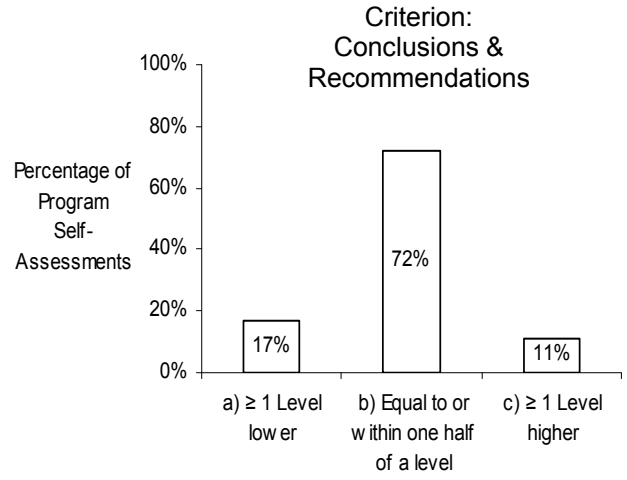


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Carlos:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **MECHANICAL ENGINEERING**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**D**	**I/E**	**E**	**D**
Primary	E-Criteria for the PLO have been articulated, assessment criteria may need further development.	E/D – Both direct and indirect evidence is included. The evidence needs further alignment with the PLO.	I/E – Faculty reviewers do not appear to have been calibrated to apply assessment criteria in a uniform way, but the FAO checks inter-rater reliability.	E/D-Results are included in the form of a rubric mapping the relevance of each course to the PLO, and the faculty allocated and student perceived commitment of time in each course to the PLO. However grade statistics are missing for the various lines of evidence. Indirect evidence is presented graphically in the form of student exit surveys.	D – Clear recommendations for improving student learning and assessment, along with discussion of the validity and reliability of the results, are included.
Secondary	I/E – criteria are identified but the PLO is very general	D – includes direct and indirect evidence, alignment is suggested by rubric evaluation	Agree	E – missing key data	E/D – student learning could not be evaluated directly and no conclusions are drawn.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A clear strength of the program is the various lines of evidence that the faculty have identified, including an accreditation database that tracks the number of hours in each course dedicated to each CLO, which in turn maps to a PLO, the student

exit surveys, mapping for course relevance to the PLO, and the intention of mapping coursework used for program assessment directly to the PLO. It is clear that the ME faculty have put into place a comprehensive framework for program assessment that they will continue to improve and build upon.

Secondary Reviewer: I agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer:

1) More evidence of communication among the faculty/instructors regarding assessment criteria, standardization of expectations, collection of course information, analysis of assessment results, and means for program/assessment improvement would be helpful. (The extent of faculty discussion was not clear in the report.)
2) As the ME FAO has articulated in the report, more direct alignment of the coursework used in assessing the program with program PLOs would be helpful.
3) A total of 15 PLOs that will need to be assessed is worrying from the standpoint of time management for assessment. Is it possible for the program to streamline the PLOs into a more manageable number, or to streamline the assessment process?

Secondary Reviewer: 4) The PLO would be more student, and generally more stakeholder, friendly and meaningful if it were revised to describe more specifically what the students will do to demonstrate “a working knowledge” and “how these principles evolve into other disciplines...” Will students demonstrate their working knowledge via, for example, the ability to solve problems or design engineering solutions to problems? To identify more specific verbs, the program might look to the questions that formed the lines of evidence used in this assessment effort or perhaps the aligned course learning outcomes.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

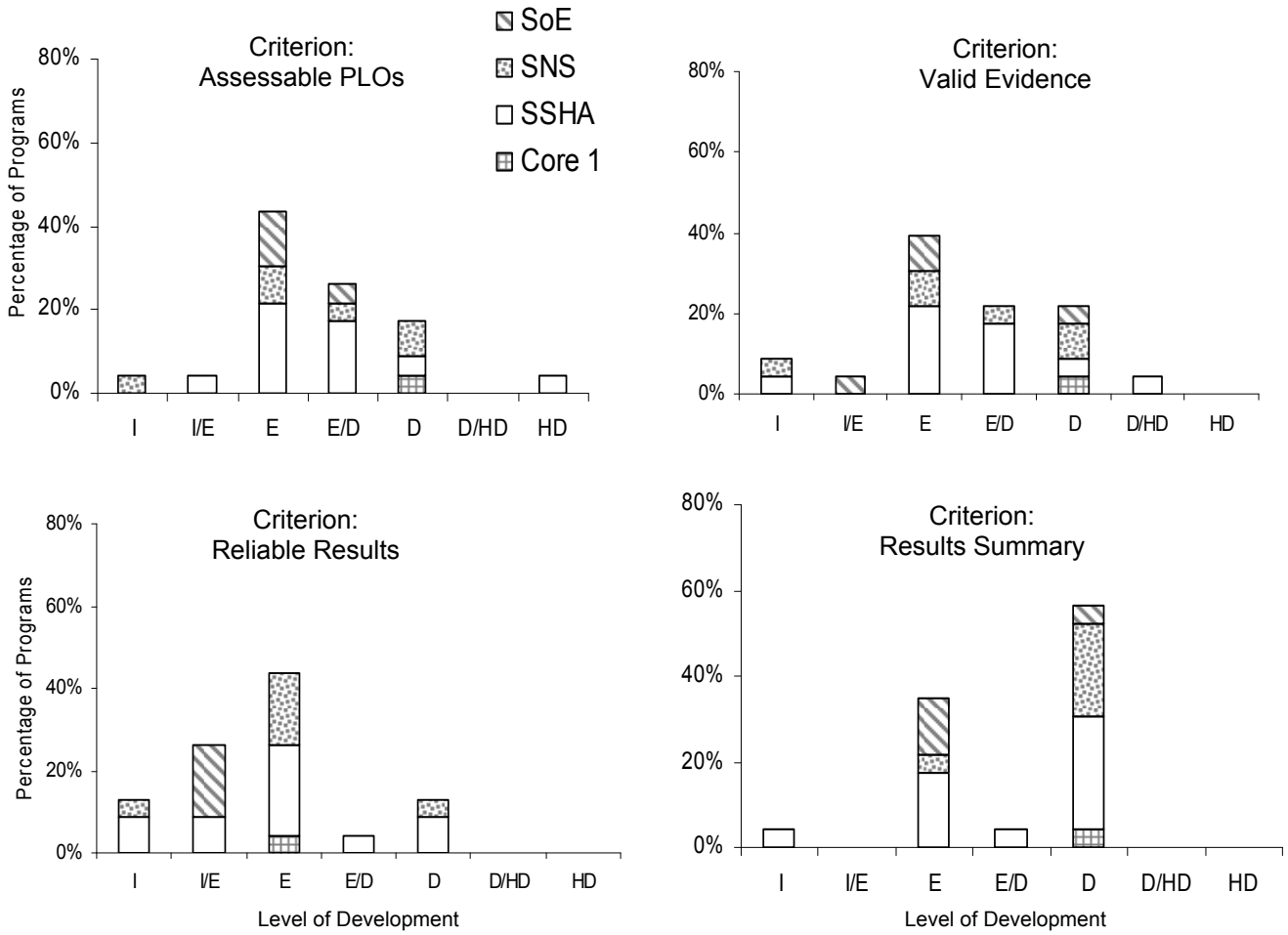
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



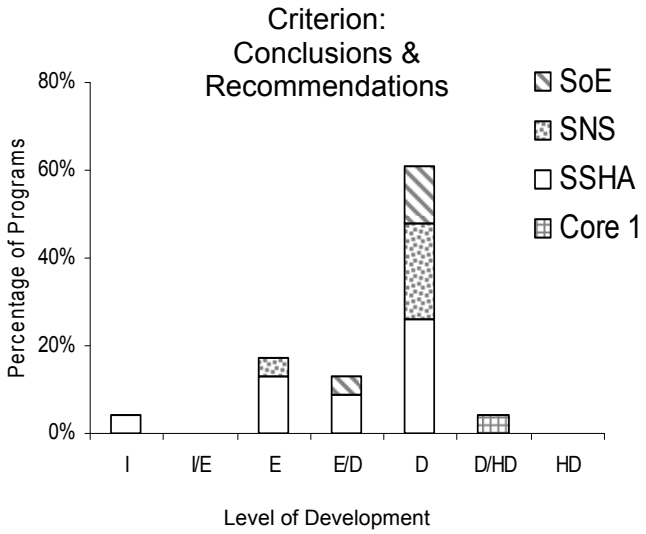


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

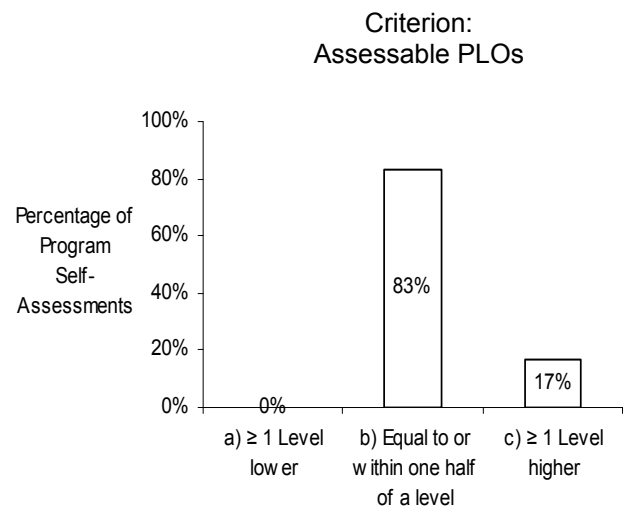
These results indicate that

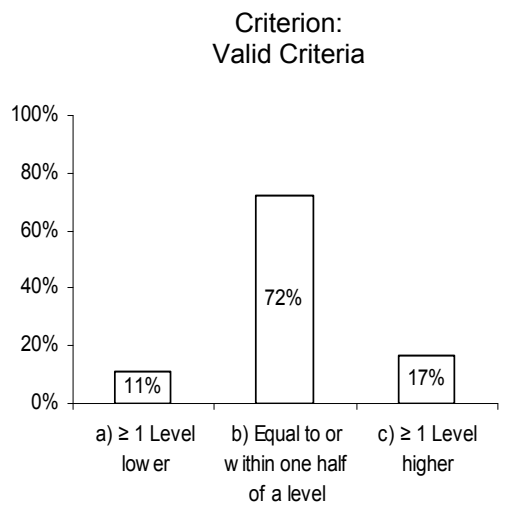
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

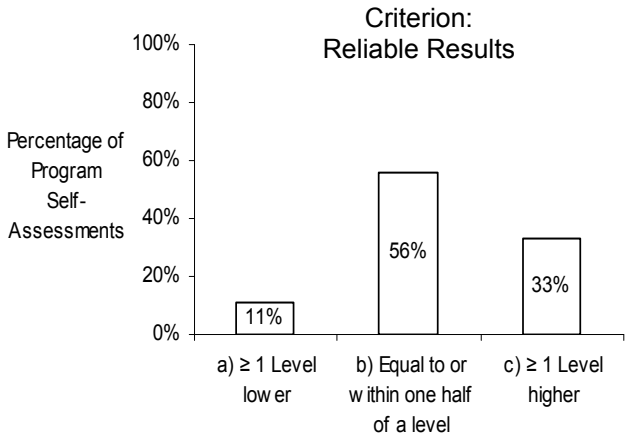
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

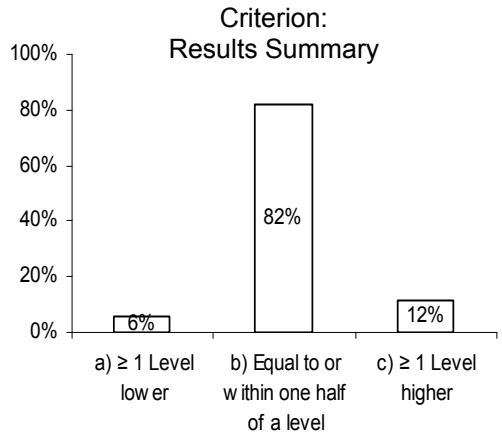
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









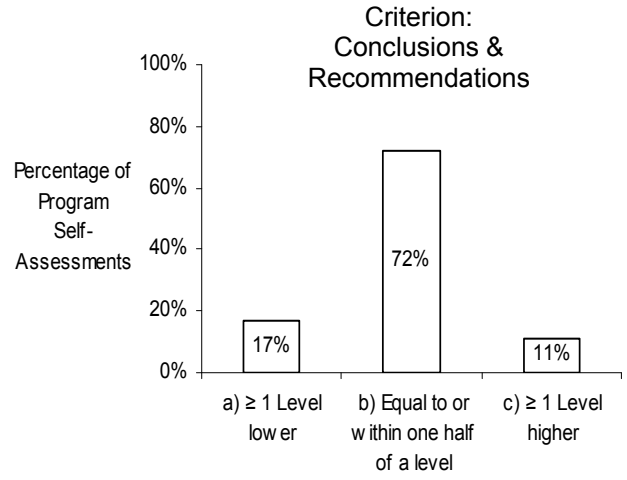


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Anne:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **MERRITT WRITING PROGRAM**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**D**	**D**	**D**	**D**
Primary	E	D	D	D	D
Secondary	E	D	D	D	D

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The high degree of faculty participation and the collaborative nature of the entire assessment enterprise as a way of improving learning. This is the real goal of assessment.

Secondary Reviewer: Agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: As the review itself notes, the PLO is not sufficiently clear to students; my greater concern is that it is focused on the wrong point. See notes below. That said, I'm impressed with how the group recognized that the rubric actually undercut consensus. Face-to-face discussion of readings seems to be required to shift the focus back to holistic grading, with the rubric merely providing guideposts, from rubric-centered, with subcategories yielding partial scores. Cf my comments on the CORE 1 assessment.

Secondary Reviewer: As the program notes, the program and students would benefit from revisions to the PLO that more clearly articulate how the students can demonstrate having engaged in the iterative process of writing with the focus on why iteration is a valuable skill to develop. Are they learning to practice it regularly and effectively? How will students demonstrate this through the portfolio?

Selecting the type of rubric, should in part, depend upon what the program wants to gain from the assessment process. While a holistic rubric and evaluation might provide the program with a sense of the overall quality of portfolio with respect to the PLO, it can make it difficult to identify particular criteria that need programmatic attention because the same holistic score can encompasses diverse combinations of strengths and weaknesses with respect to the underlying criteria.

4) Any outstanding thoughts or questions?

Primary Reviewer: The wording of the PLO bothers me: "students will be able to demonstrate thorough engagement with the iterative processes of reading, writing, and speaking." It may accurately reflect the assignments given, but I'm not sure it asks if students have in fact learned how to make these iterative processes their own. In other words, shouldn't it be "students will engage the iterative processes of reading, writing, and speaking as fundamental learning tools"? Or something like that. In other words, what's this iteration supposed to do? What is it *for?* Maybe, "Students will engage the iterative processes of reading, writing, and speaking in order to deepen their understanding of complex ideas? In order to clarify and improve their writing and speaking? Maybe the Outcome just needs a preamble to explain why we want to engage in these iterative processes, and that would enable the program to leave the PLO more or less alone (though I'd still shift "engagement" to its verb form).

Secondary Reviewer: The possibility that the uniform appearance of electronic portfolios (relative to paper versions) undermined the ability of raters to assign low scores, even when the written work was clearly inferior, is intriguing. This speaks to the diversity of cues that faculty, perhaps subconsciously, apply in determining or perhaps rationalizing their scoring and how the template-bound, electronic approach might help to focus evaluation and the associated discussions.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

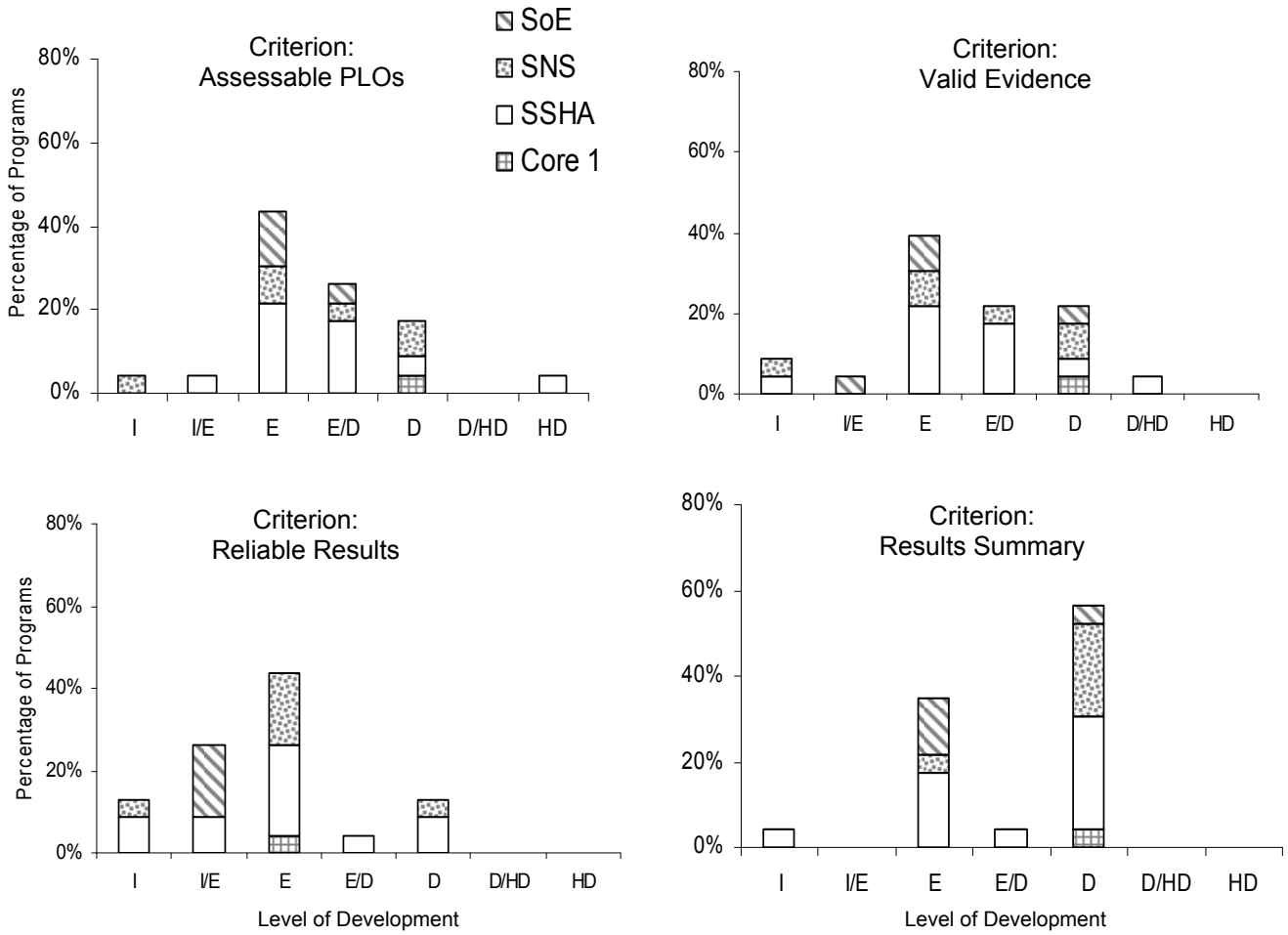
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



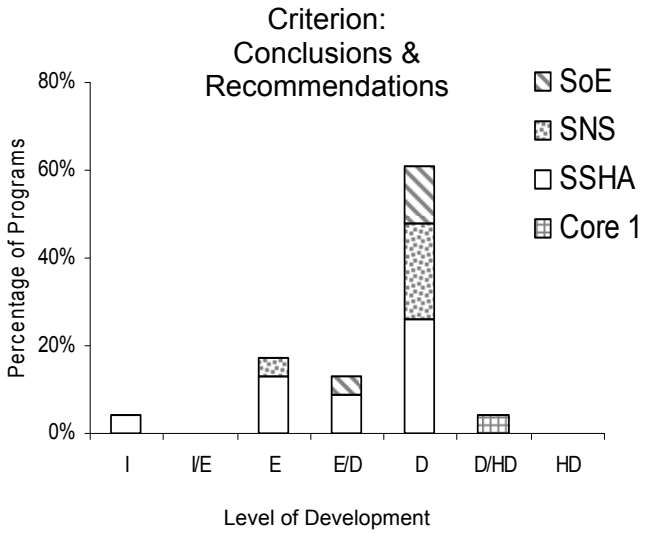


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

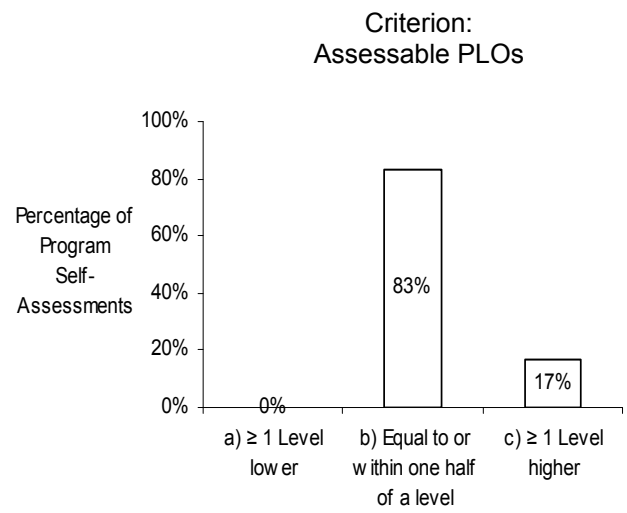
These results indicate that

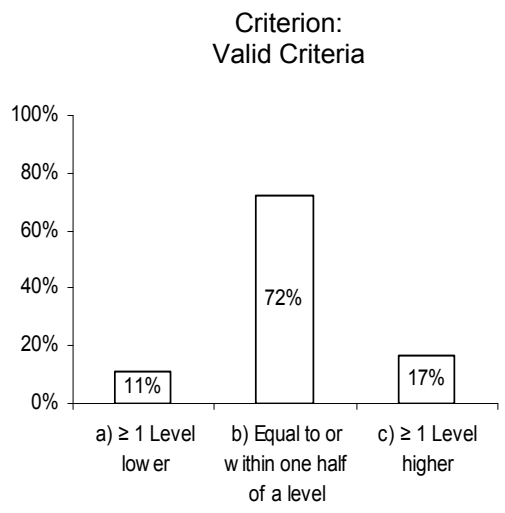
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

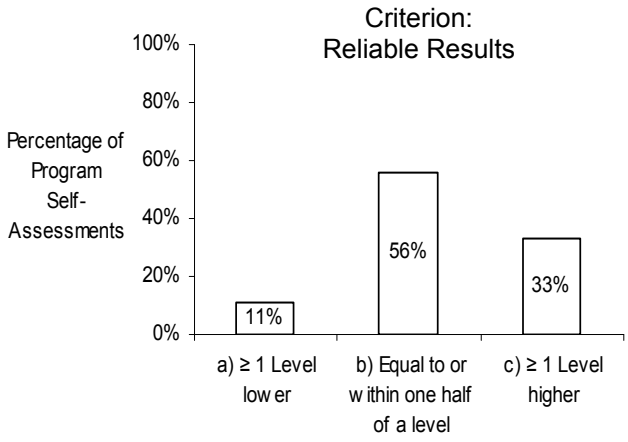
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

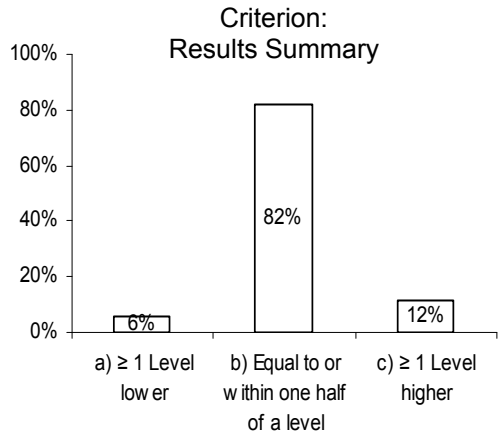
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









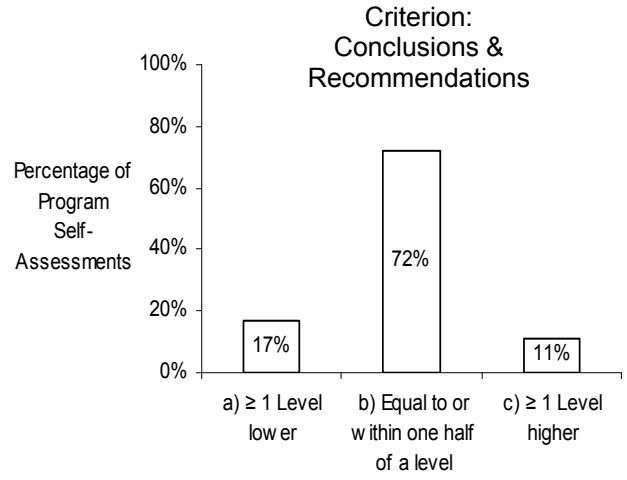


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Mayya:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **NATURAL SCIENCES EDUCATION MINOR**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**D**	**D**	**E**	**D**	**D**
Primary	D – have reasonably assessable outcomes that include criteria that identify key elements of the outcome. No rubric but not how they plan to assess. "Be familiar' doesn't indicate how students can demonstrate learning. No standards identified.	D- direct and indirect evidence; careful attention to alignment although not perfectly aligned with, for example, PLO 1 which expects students to be able to "comprehensively articulate" but they aren't given this opportunity. Results however are meaningful.	E- but not really applicable in the sense that reliability addresses the ability to use a multiple choice answer key. No information is given about how open ended questions responses are scored.	D – clearly summarize results. No benchmarks.	D – summary & analysis could have been enriched by looking at points of intersection between direct and indirect evidence. For example, how do students self-perceptions align with their performance on these assessments?
Secondary	E/D – No rubric for the assessment is developed.	E/D – both direct and indirect evidence were collected, but the evidence of faculty agreement on the types of assessments was not clearly stated.	E – Not clear about number of faculty who were involved in assessing evidence.	D – Results clearly delineated for each line of evidence and are summarized.	D – Clearly articulates conclusions, implications and recommendations for improvement.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A real strength of the program's assessment work is the careful attention and integration of course with program-level goals and the comprehensiveness of these efforts across the program, including the effort to align assessment questions with the PLOs.

Secondary Reviewer: Agree.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) One clear benefit of the current assessment approach is the efficiency of multiple choice scoring, permitting assessment across the program with limited human resources. Some of the questions, however, don't tightly align with some of the skills suggested by the PLOs, for example, "comprehensively articulate" or "develop ..and deliver" or "design". Do students practice some of these activities in the classes offered (at UCM not in the field which is recognized to be a more complex venue for assessment) and would it possible to develop a rubric to evaluate some of these key skills? Put another way, perhaps there are pre-existing opportunities in the extant curriculum to examine the actual implementation of some of these skills? (2) Related to 1, one benefit of having such tightly aligned indirect and direct assessment questions is that the program can directly compare student learning results (the multiple choice questions) to the student's perceptions of their learning. Doing so, might identify informative differences in the proficiencies demonstrated by the questions and the students' perceptions of their learning that might be interesting for the program to pursue for further alignment.

Secondary Reviewer: Agree.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

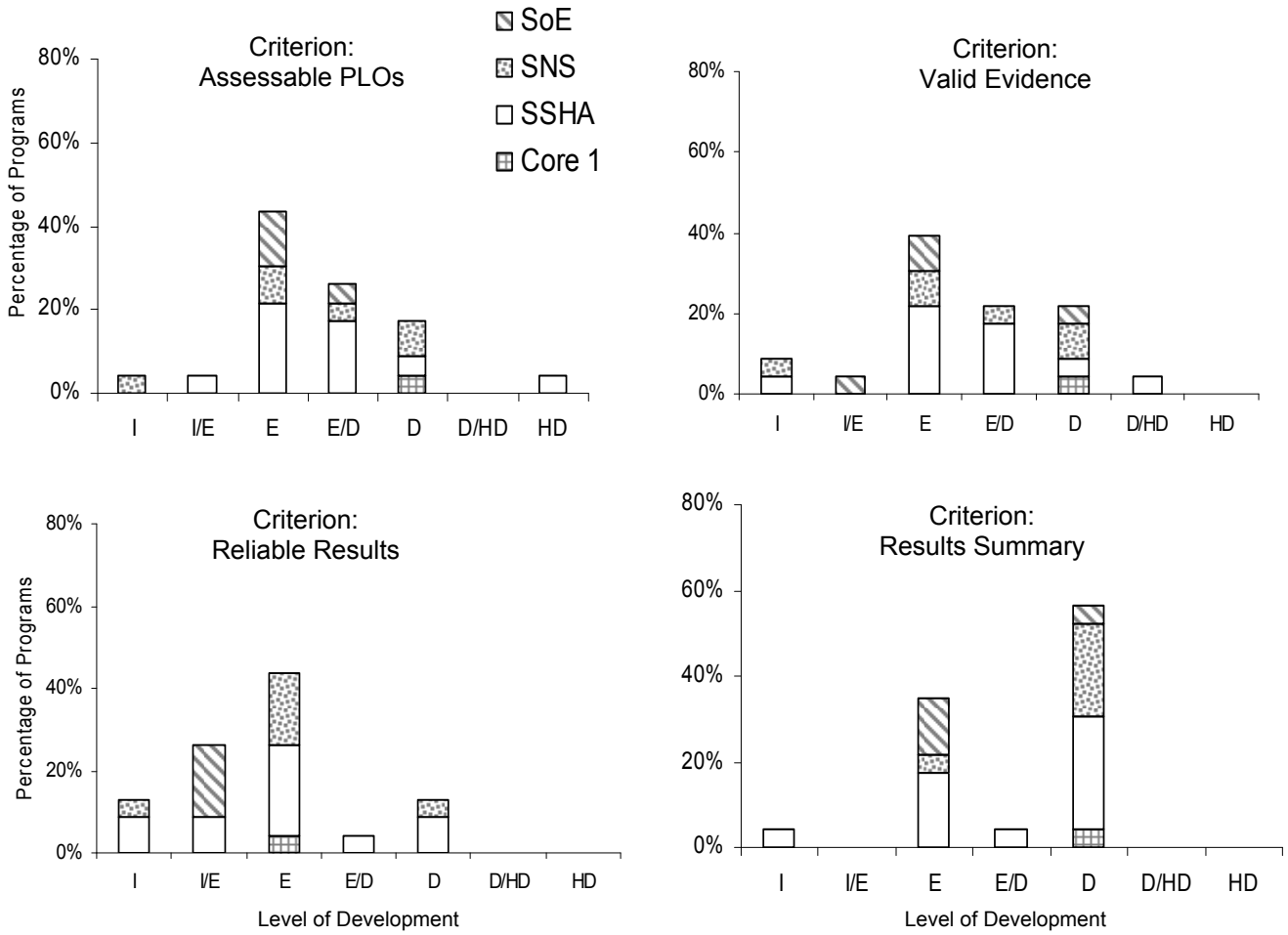
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



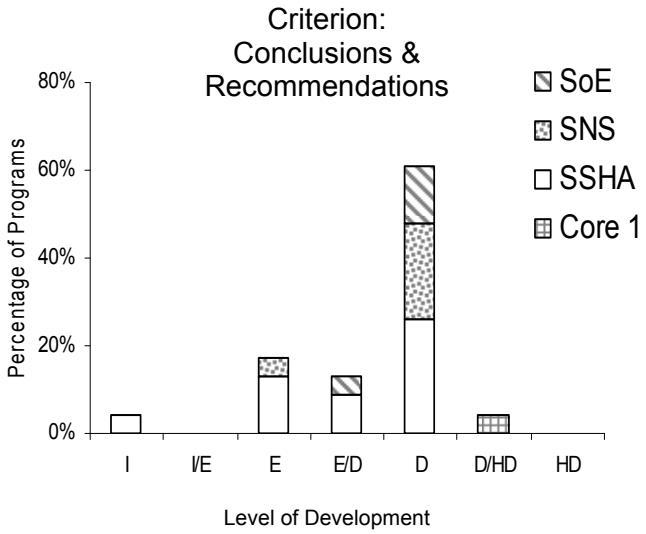


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

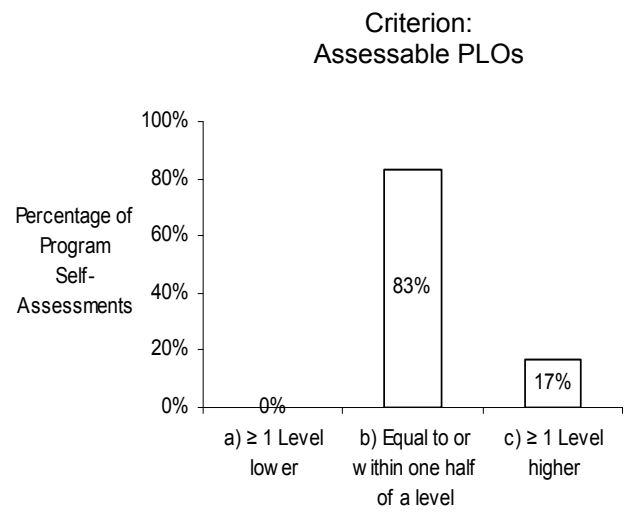
These results indicate that

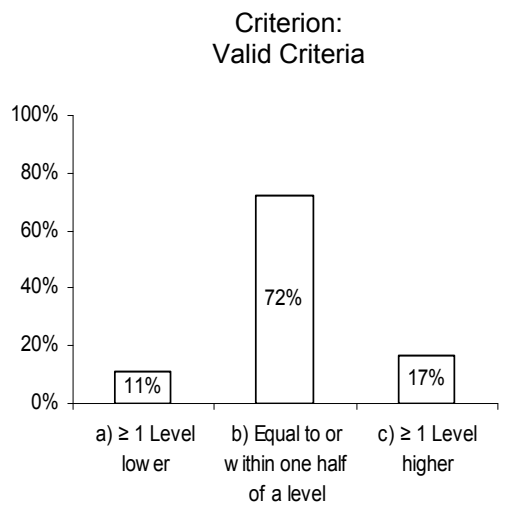
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

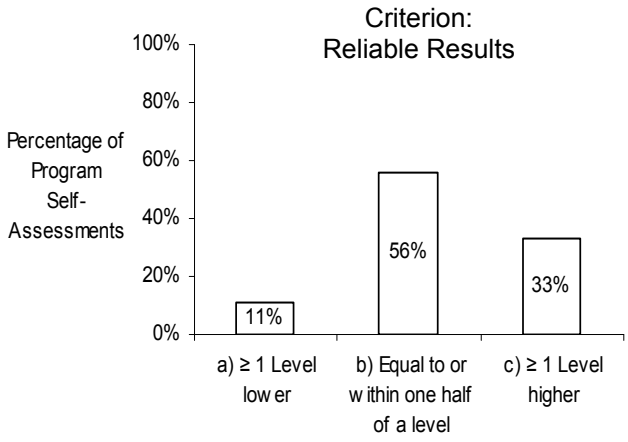
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

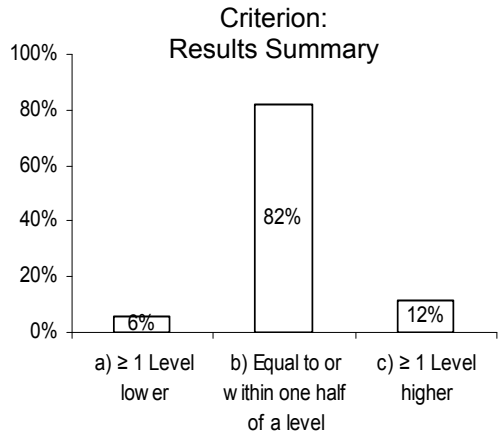
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









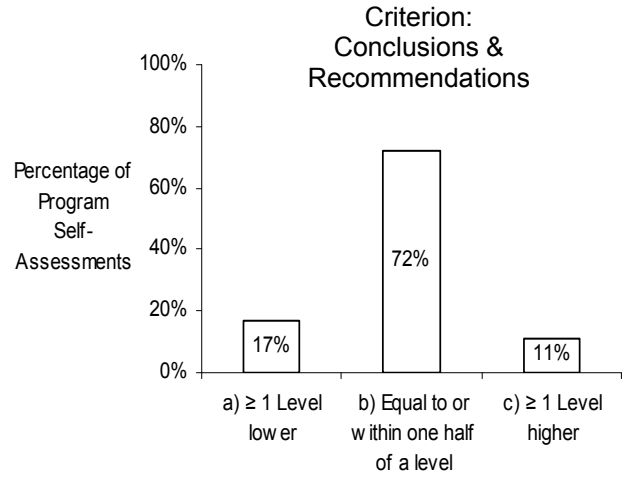


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Peter:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **PHILOSOPHY**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E/D**	**E/D**	**E**	**E**	**E**
Primary	E/D. Use of embedded items that directly tap the very specific skills described in the PLO is a strength. However, without actually seeing the items used and the rubric by which they are assessed, it is difficult to judge whether the rubric is emerging of developed. Including that information in the PLO report would be a relatively simple change.	E/D. Only direct evidence is considered. The nature of the direct evidence does not necessarily support the sort of longitudinal tracking of performance that is ultimately hoped for, because no evidence is provided that the difficulty of items used in different years is equated.	E. It is unclear who scored the items that were used to assess the PLO. Given that these were questions embedded in the final exams of a single instructor, it seems plausible that he was the sole scorer. If that is the case, there is no check on scoring reliability. This may not be a problem if the scoring rubric is sufficiently clear and objective, but in the absence of that rubric, the PLO assessment does not provide evidence of reliability.	E. As the PLO report points out, what can be done in the summary of results is necessarily limited by the small sample size.	E. Although the validity of the assessment appears to be good, there is insufficient attention to reliability.
Secondary	Agree	Agree	I/E –The assessment criteria appear to have been unique to a single faculty member. The goal is to develop shared/programmatic expectations of student performance.	E – It is difficult to interpret the figures as there are no legends and I don't know what the quantities on the axes represent. I think they are frequency histograms but with the frequencies on the X axis.	Agree

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The assessment employs what are probably highly objective measures of very well defined skills.

Secondary Reviewer: The program has clearly and precisely defined learning outcomes and a clearly defined rationale for selecting this course as a source of evidence for assessment of these outcomes.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: The biggest shortcoming of this effort is the questionable equitability of questions that, though tapping the same skills, differ in specific details (and hence potentially difficulty) across years. It might be worth considering identifying a set of embedded questions that will remain constant.

Secondary Reviewer: 1) The program needs to continue to develop a program approach to assessment. 2) This would be supported by including in the report examples of the student work demonstrating different levels of proficiency identified in the report and the associated scoring rubric. Attending to this sort of record keeping and information sharing will be particularly important as the program grows to include more faculty. 3) Additionally, since PLOs are intended to represent the skills and knowledge graduates demonstrably possess, assessment would also be more programmatic if evidence was included from more than one class. Examining student work from an upper level class, in addition to this introductory course, will help the program determine what skill level most graduates should exhibit, and if this core course and the program more broadly is providing sufficient opportunities for development of this proficiency. 4) It will be important for the faculty to integrate indirect assessment into their assessment process. Since the number of graduates is low, the faculty could consider exit interviews or focus groups facilitated by the SATAL program.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

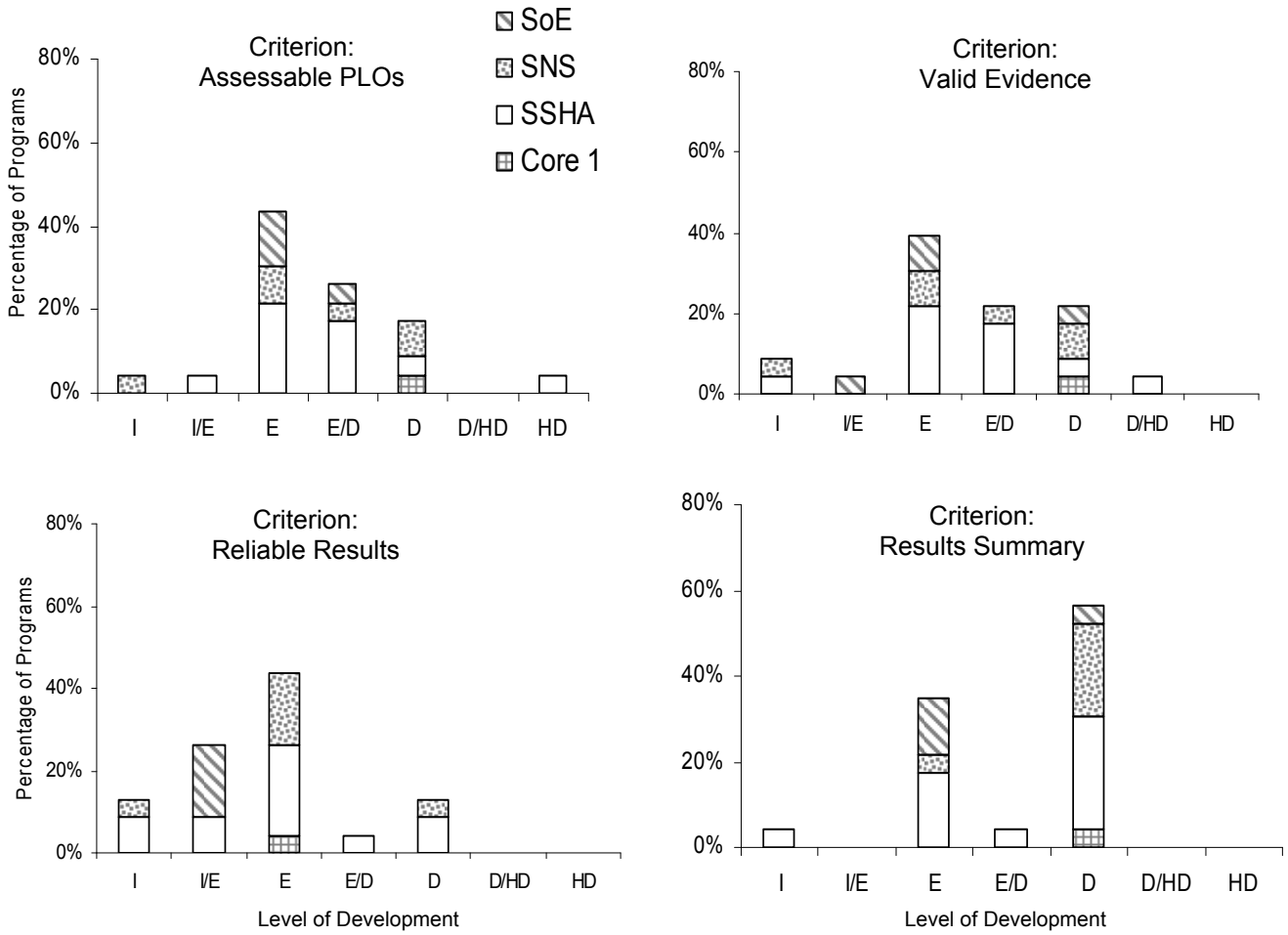
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



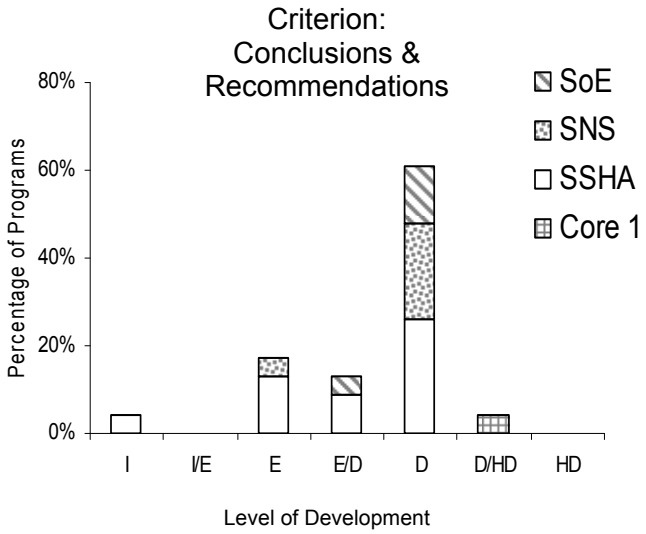


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

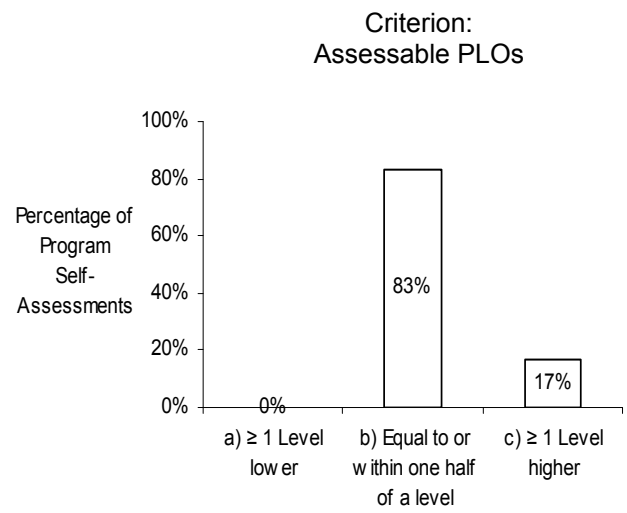
These results indicate that

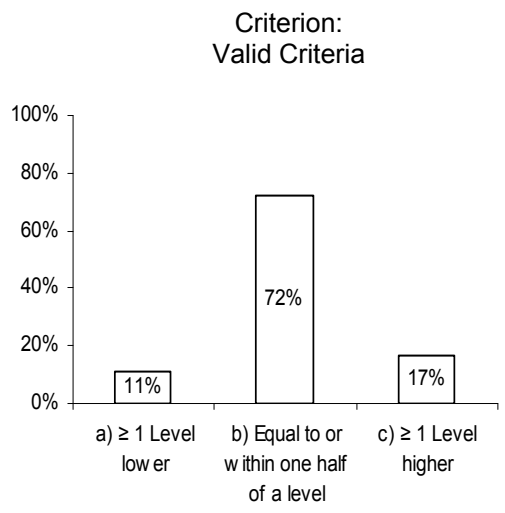
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

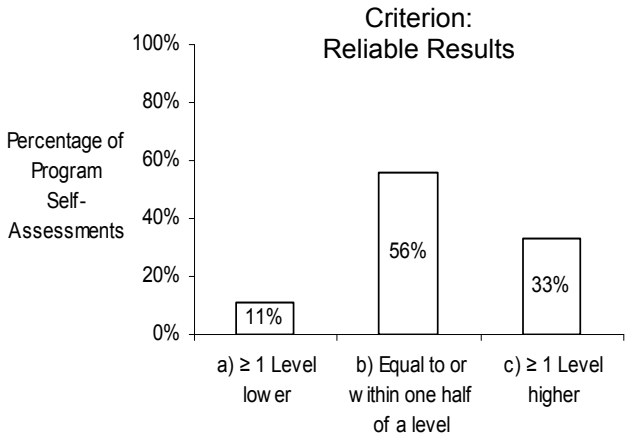
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

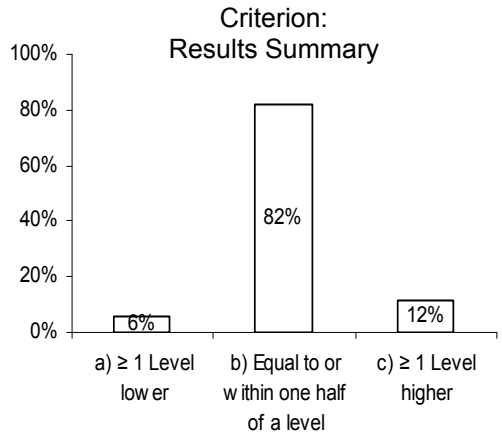
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









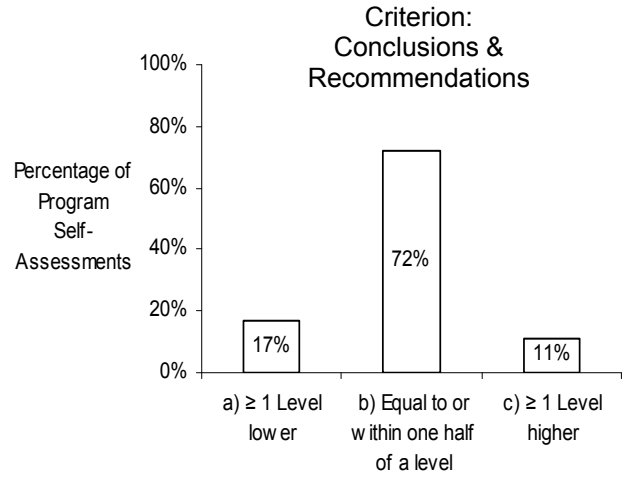


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Kevin:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **PHYSICS**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**D**	**D**	**E**	**D**	**D**
Primary	D – have reasonably assessable outcomes. The rubric is defined, but required more development of assessment criteria.	D – Collected relevant and sufficient direct evidence for the PLO. Indirect evidence from student focus group were collected, but not clearly aligned with the PLO.	E/D – Closer to D. some faculty calibration was done, but not for all reviewers.	D – Results clearly delineated for each line of evidence, but wish they had also included summary of student focus group analysis.	D – Report clearly articulates conclusion, implications and recommendations for improvement. They may need to discuss how physics labs will be assessed against the PLO.
Secondary	agree	Agree but wonder too about the alignment of the direct evidence with the PLO – did the questions cover all three skills the PLO highlights?	E – the report is not clear about faculty calibration but does include some but not all reliability results	agree	Agree – my only hesitation is the small size of their focus group and taking action based on that.

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A real strength of the program's assessment work is that many faculty and lecturers in the Physics program took careful consideration of the PLO and comprehensiveness of their efforts across the program, including the careful alignment of questions with the PLO.

Secondary Reviewer: I agree with the first statement and would add that the program has implemented all aspects of program-level assessment – broad faculty involvement with evidence collected from multiple courses. Their assessment protocols were also sufficiently well developed and conscientiously applied so as to produce meaningful and actionable results.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

> *Primary Reviewer:* (1) How are the outcomes of the student focus group assessed? What were conclusions and recommendations by Student focus group? The outcomes of this group can be also used to develop better curriculum and physics pedagogy.
>
> *Secondary Reviewer*: (1) As physics noted, I would encourage reporting all inter-rater reliability results to promote confidence in the conclusions. (2) It would be helpful to have some sense of how the questions examined mapped on to the three skills Physics wants to see in their graduates – the ability "to explain, analyze and predict." Each skill is distinct, engaging physics knowledge in different ways, and in doing so requires the development of different neural pathways for engaging the same content knowledge.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

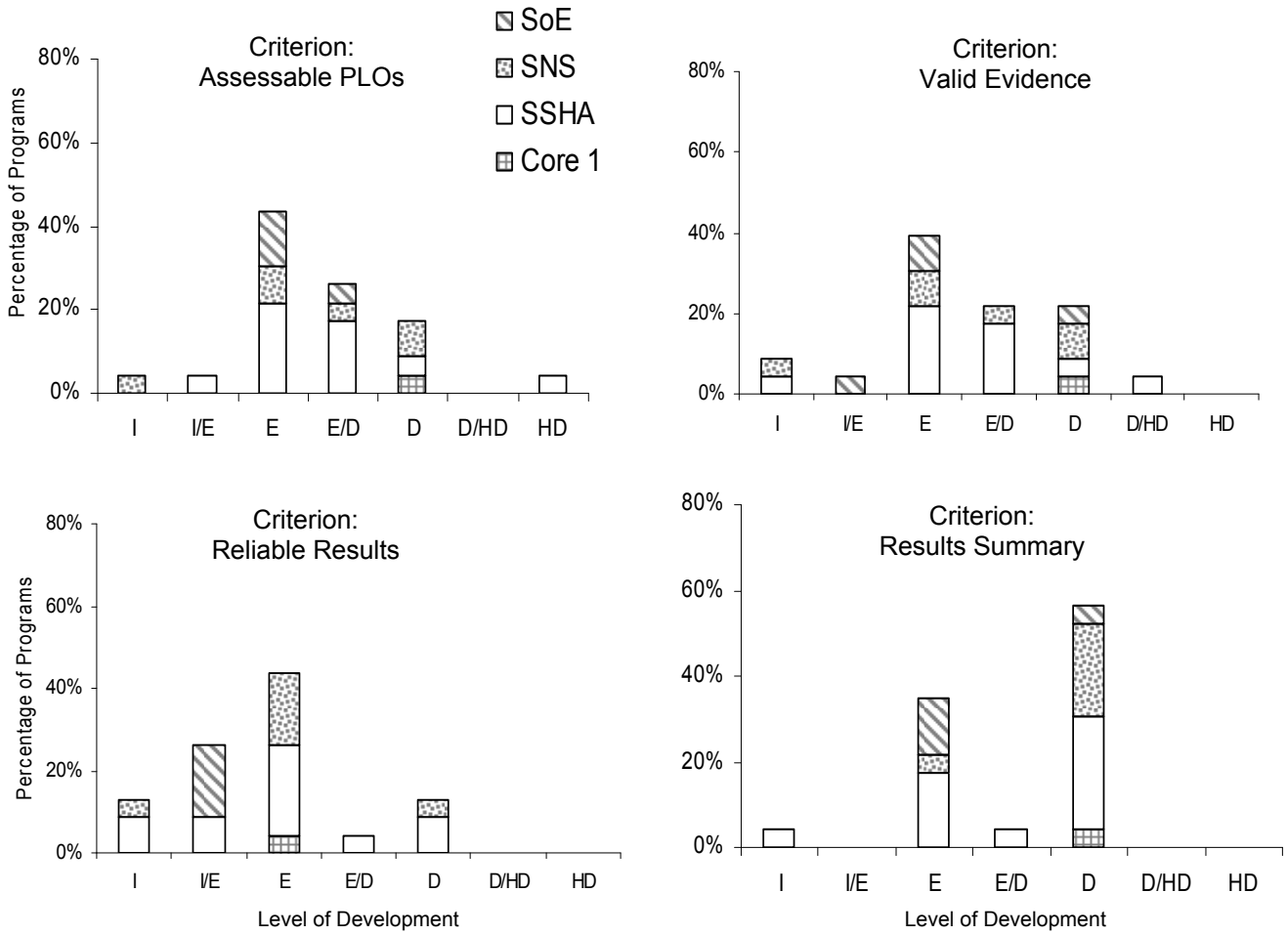
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



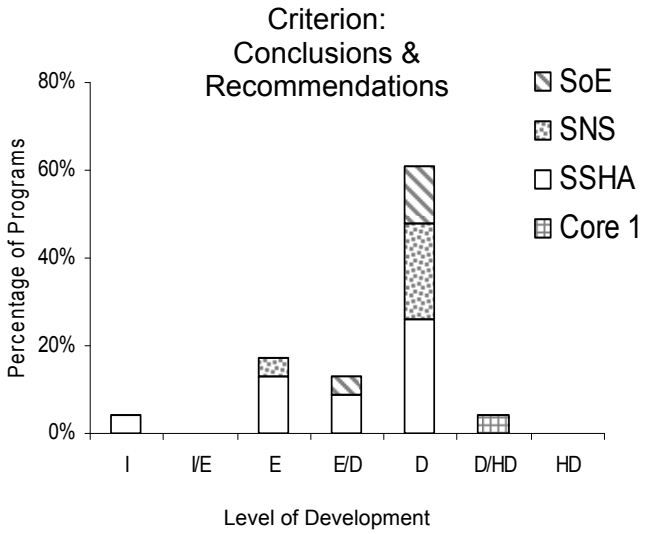


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

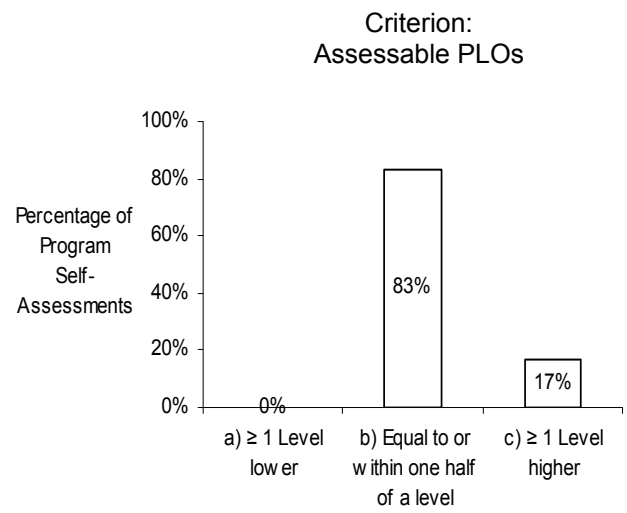
These results indicate that

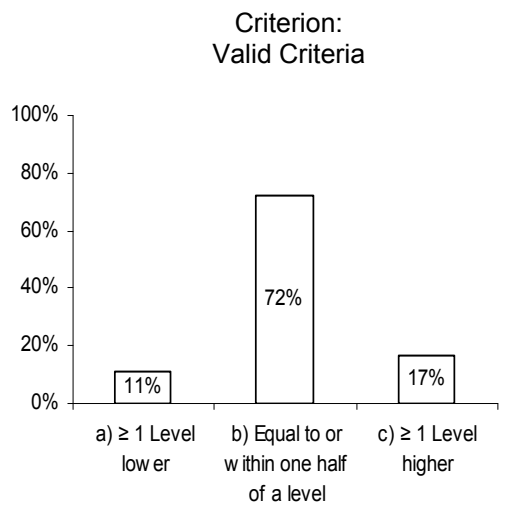
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

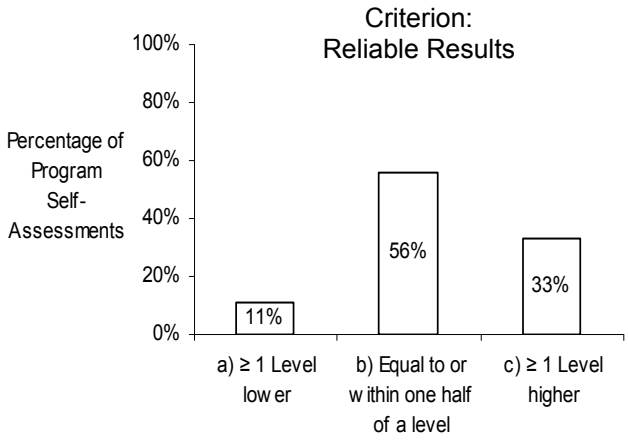
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

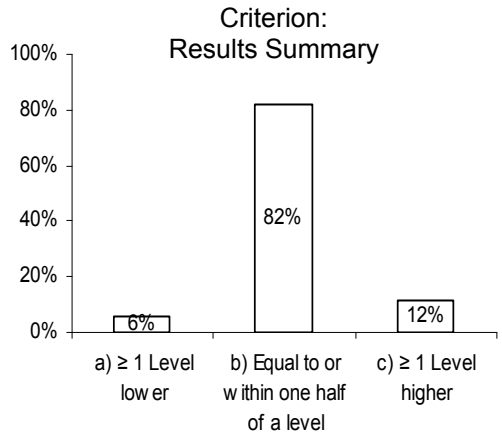
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









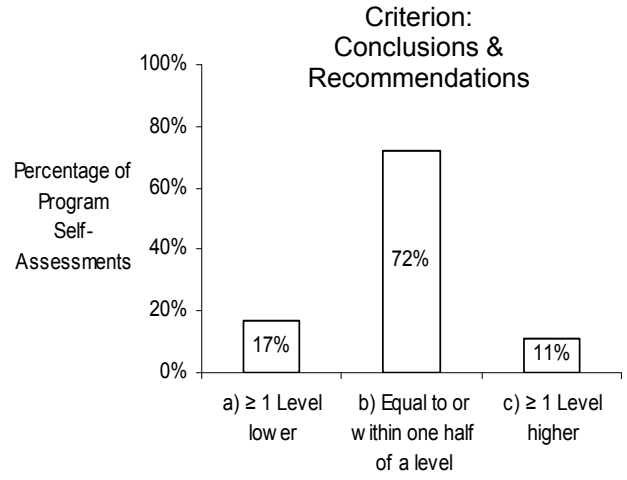


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Nate:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **Political Science**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Committee	I/E – An informal rubric exists and the PLO needs further development to be measurable.	E – vagueness in report makes it difficult to ascertain if program or course-level assessment.	I/E – reviewers seemed to engage in informal calibration but need to formalize and gather evidence of how often raters came to the same conclusion (inter-rater reliability)	E – have some idea of student proficiency as student work examples helpful in this way. Did not report the sample size and the frequency of different levels of student performance.	E/D – Lack of detail hinders conclusions

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A clear strength of the program's assessment efforts is the successful identification of a weakness and the identification of a plan to adjust the program in response.

Secondary Reviewer: The attention paid to the alignment between the PLO and the structure of the embedded question was exemplary.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

(1) Continue to further develop an inclusive, programmatic approach to assessment by adopting several practices including: develop a rubric that describes the criteria and standards that emerge from this first round of assessment and are addressed in this report, revise the PLO to include a verb(s) that specifies how students will demonstrate this learning/understanding, and outline more formally the focus group procedure and questions. Keeping an explicit record of this work will be important as the faculty grows and will help current faculty to interpret and learn from the work reported here. 2) As noted in the PLO Report itself, there is a need to make the assessment more representative of student learning as a result of the program as a whole.

Consider examining student work from more than one course. 3) As was done for the focus groups, please describe the student population that was assessed with the embedded questions, including key attributes like the sample size, the students' degree of academic advancement (ex. junior or senior, etc.), and the proportion that are majors or minors as relevant. This will help the faculty to interpret the significance of the results of the assessment in terms of expected student performance in the trajectory of the major. What students achieved what levels of proficiency and what are the implications for developing student learning through the major?

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

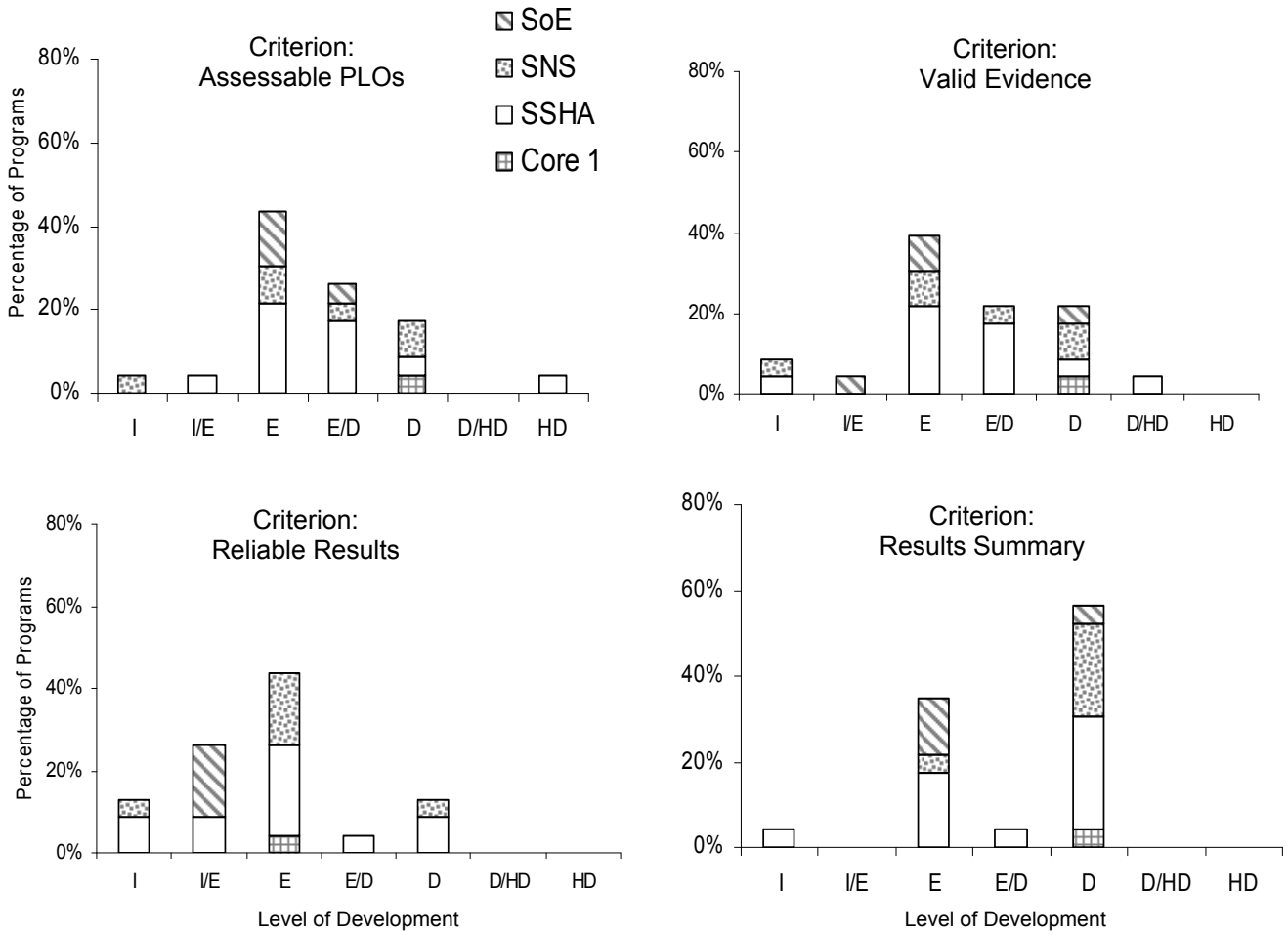
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



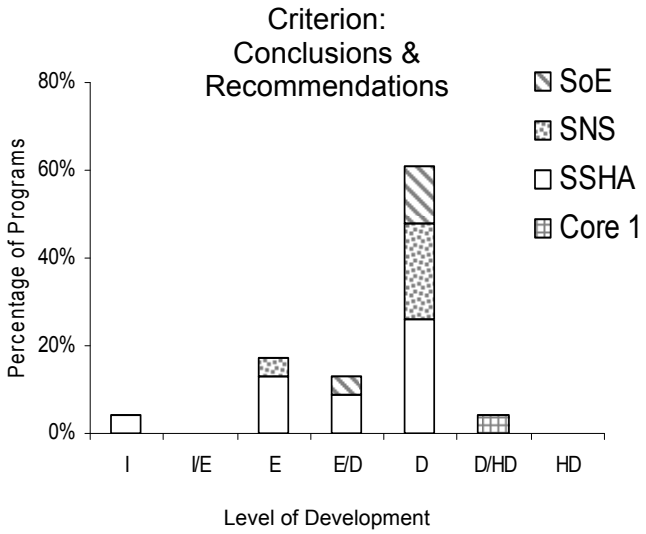


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

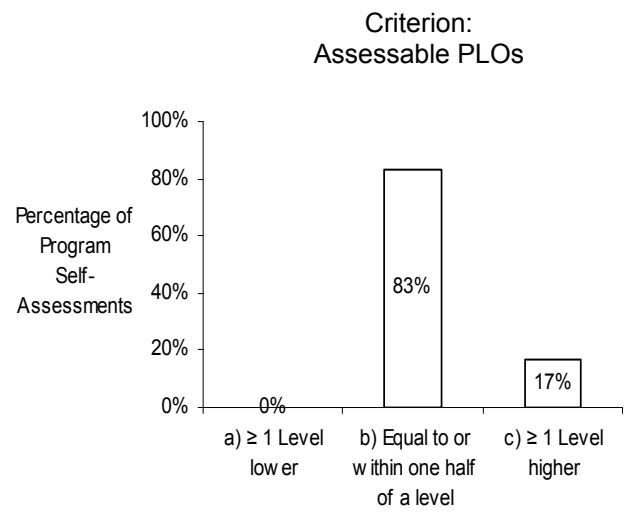
These results indicate that

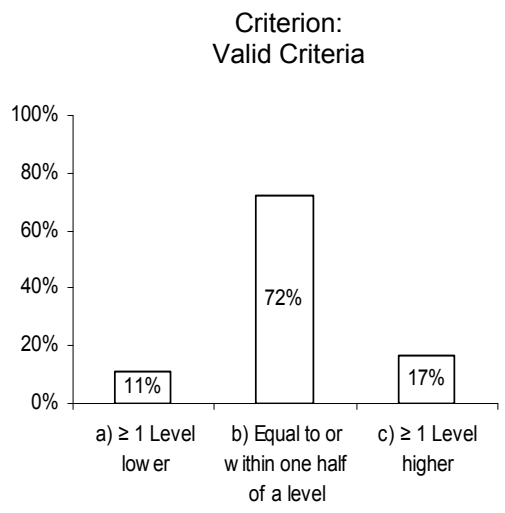
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

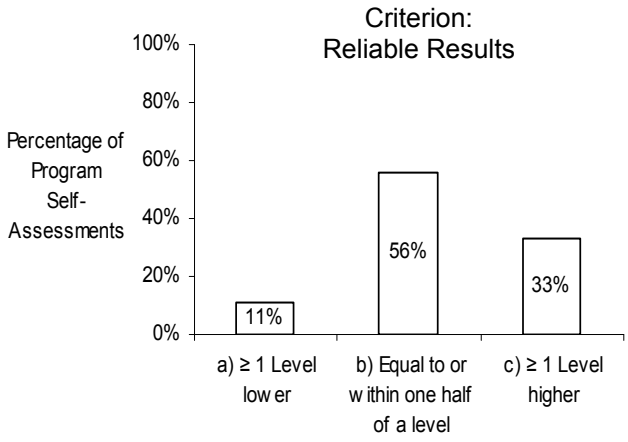
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

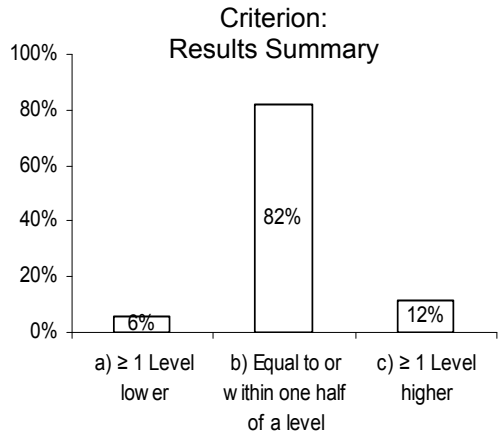
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









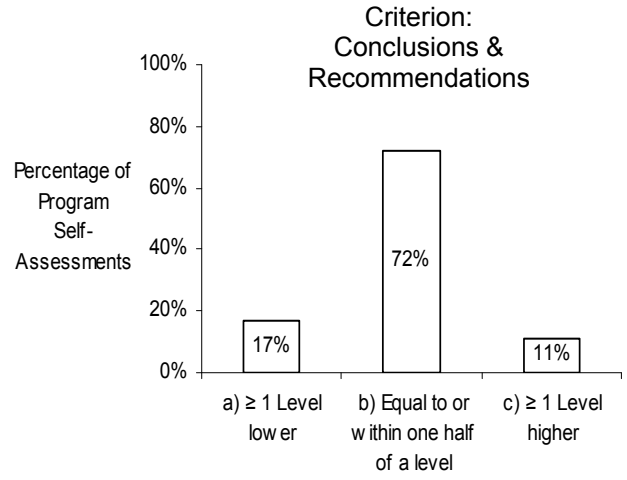


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

U N I V E R S I T Y O F C A L I F O R N I A

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Will:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **PSYCHOLOGY**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E**	**E/D**	**E/D**	**D**	**D**
Primary	**E** (Though it covers the range of knowledge in psychology, the PLO's scope may make it difficult to assess all at once, and its action verb can be more active and specific to describe how students can demonstrate learning [as is noted in Self Evaluation section.)	**E/D** (The ETS major test appears to represent the faculty's collective sense of the discipline, and thus provides relevant and consistent evidence for the PLO being measured. The sample size is small and may not provide sufficient program-wide data. There might be more of an intensive rationale for using the test for this assessment.)	**E** (Although the ETS major test appears to represent the faculty's collective sense of the discipline, its use doesn't allow for faculty engagement in the process of assessment save for interpreting results and collecting indirect evidence.)	**D** (Results are clearly reported in tabular and comparative form, also pertaining to each portion of the PLO, although the specific supporting student work—such as actual test questions and student answers—could be included in some form as well [as is suggested in "Conclusions and Recommendations.")	**D** (There is extensive consideration of the aptness of the ETS major test and the program's role in preparing students for it. There is also useful consideration of implications for program definition and future assessment.)
Secondary	E	E/D	D—The advantage, though, is the presence of disciplinary benchmarks, and interpretation of this data by such benchmarks can be very useful.	D	D

2) A clear strength of the program's assessment efforts.

Primary Reviewer: The program is actively engaged in determining its students' standing comparatively (i.e., with respect to other Psychology programs), and in examining the extents to which it could/should adopt curricula or benchmarks of other programs in defining itself.

Secondary Reviewer: Agreed.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: The use of an external standardized test may not allow for active program involvement in the assessment process, as it leaves the determination of method and results to external authorities. In addition to using the ETS major test for the purpose of assessing program PLOs, there might be other means of assessing UCM psychology students more directly and specifically. For instance, to help resolve the value of this approach with respect to the information gained from benchmarking, it might be useful to compare the student preparation goals and/or relatively unique aspects of UCM's program to those of other programs that use the test.

Secondary Reviewer: Continue to weigh the benefits and costs of this approach to assessment, including the alignment between the test and the program's curriculum and the challenges in gaining participation. To help resolve the value of this approach with respect to the information gained from benchmarking, it might be useful to compare the student preparation goals of UCM's program to those of other programs that use the test. Are they the same as this program, which would increase the value of such a comparison?

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

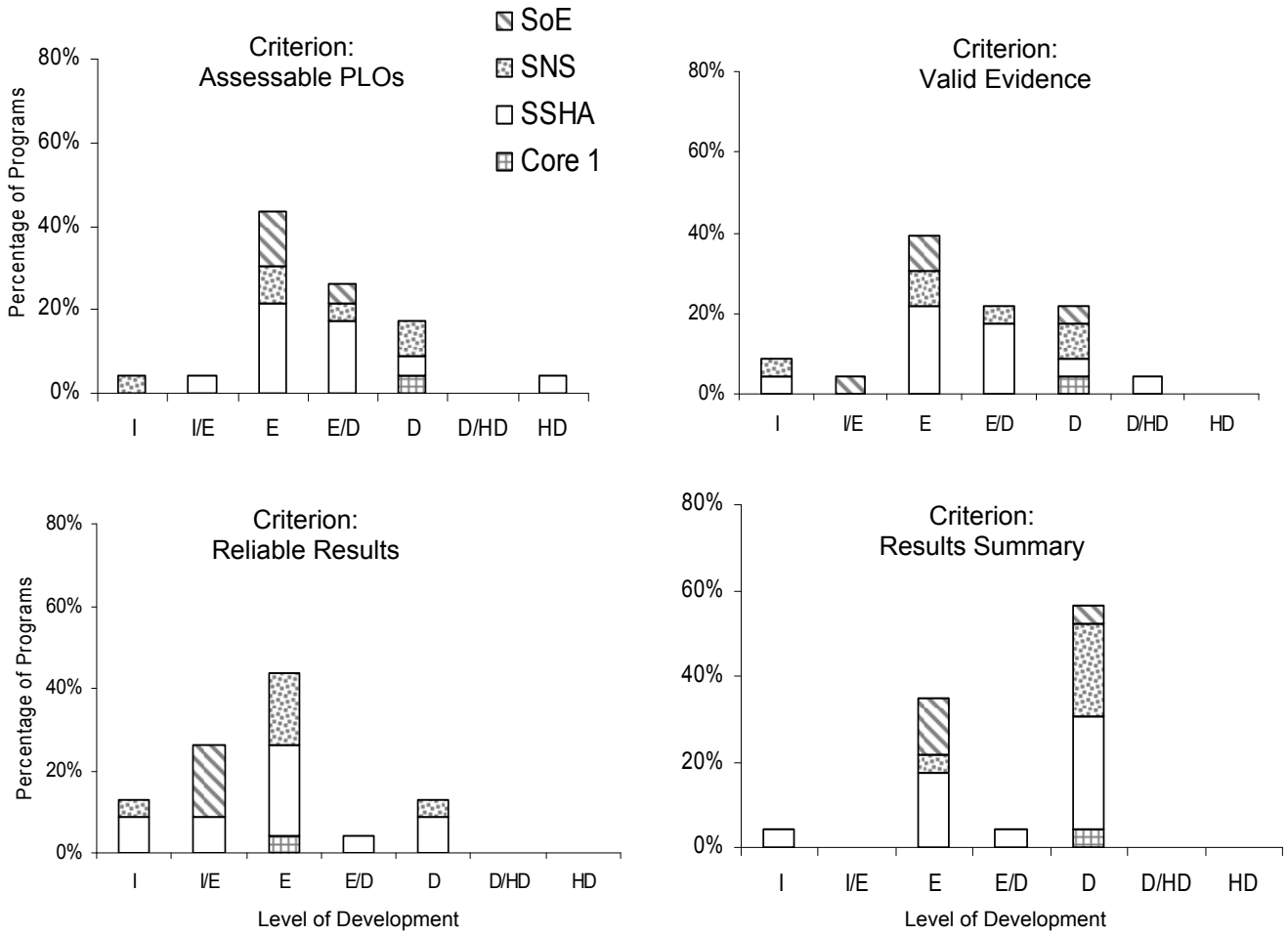
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



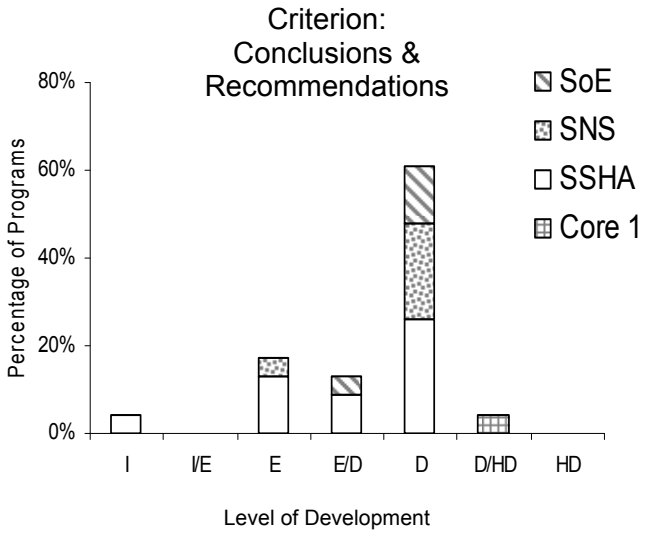


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

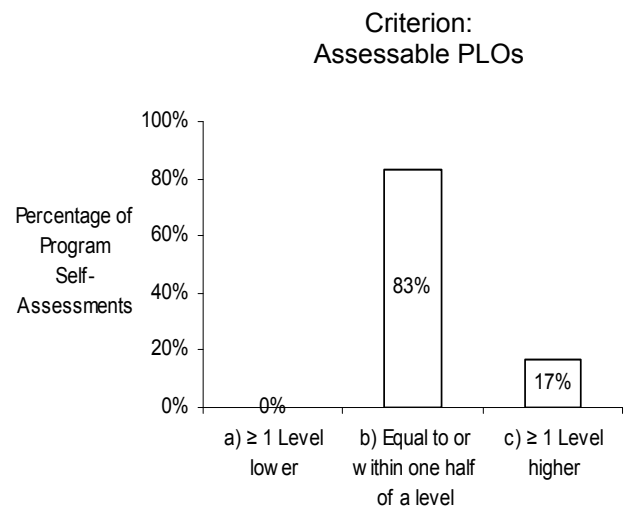
These results indicate that

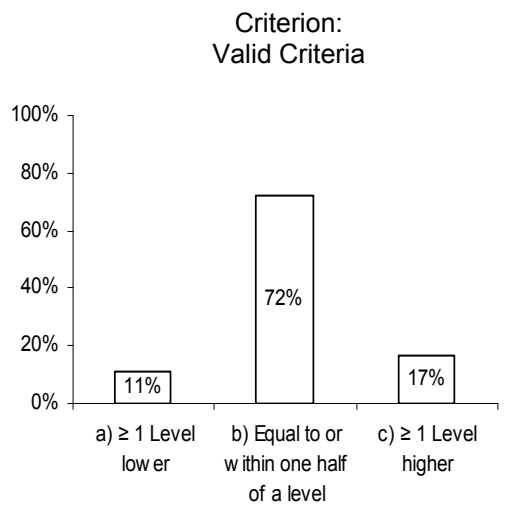
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

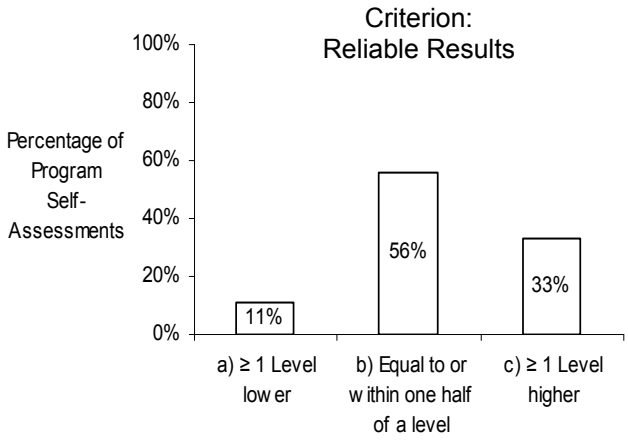
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

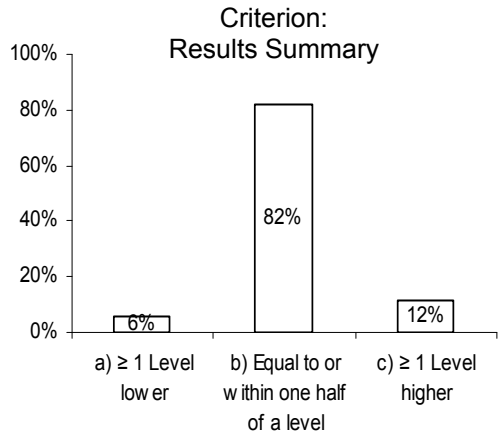
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









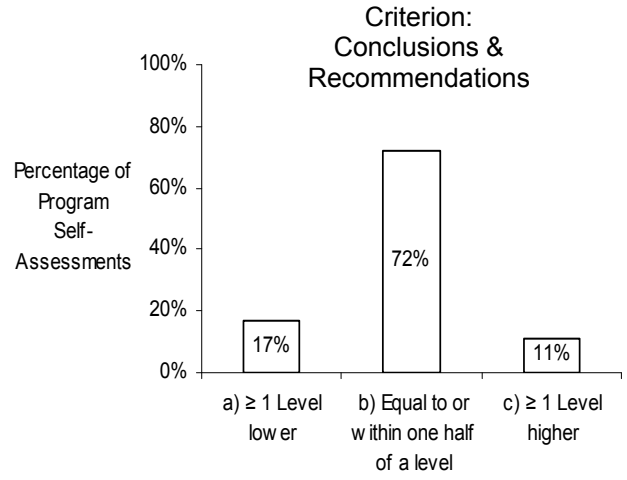


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Nella:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **SOCIOLOGY**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**E/D**	**E**	**I/E**	**D**	**D**
Primary	E. PLO #1 ("Think critically") begs for an empirical definition. How can students demonstrate successful critical thinking? (The other four PLOs focus on more directly observable actions.)	E. There appears to be no use of indirect evidence.	I/E. Although there is a clear rubric used by all raters that is intended to promote inter-rater reliability, there is no mention of a process to monitor that reliability.	E/D. The results are clearly presented in graphic form, and the split employed in Figure 2 was insightful. Although four aspects of students' written work were rated, only two were discussed in the report. It is somewhat frustrating that the writing ratings were not more fully discussed, as the mean rating (near the 'good' level) appears somewhat inconsistent with the writing samples that are provided.	E/D. The report is clear about the faculty's interpretations and plans for improvement. A plan to change one aspect of the minor has emerged directly from this assessment. However, there is insufficient attention to reliability and validity of the evidence.
Secondary	D – developed; the PLO assessed was "the ability to design and evaluate empirical sociological research." They have a rubric, with criteria and standards defining key aspects of these abilities	E/D – leaning more toward D; but not completely D because of lack of indirect evidence. Tight alignment among PLO and assessment work and rubric.	Agree; the program concludes there was high inter-rater reliability but we cannot judge that as the data are not shared.	D – I don't disagree with the primary reviewer except that perhaps it would have been important for Sociology to explain more clearly why they only focused on those two criteria. I think it is because they	D

				are the only two criteria that are directly related to the PLO examined. So maybe the point is to explain the reason for including the other two criteria, which might just be because provided additional insight.	

2) A clear strength of the program's assessment efforts.

Primary Reviewer: There is a clear rubric for the assessment of students' written work.

Secondary Reviewer: The strengths are: the involvement of all faculty in this process, the tightness of the alignment between the PLO, student evidence and rubric, and connection to the curricular changes implemented by the program.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: Currently, the program is small and all three faculty rated all written work. As the program grows, that will no longer be feasible, and the importance of assessing rater reliability will become even greater. The faculty should consider altering the first PLO to focus on something more directly observable.

Secondary Reviewer: 3) It will be important for the faculty to integrate indirect assessment into their assessment process. In the short term, when the number of graduates is low, the faculty could consider exit interviews or focus groups facilitated by the SATAL program. In the longer term, they might consider Anthropology's strategy, developing a set of questions that are administered to their students as part of the graduating senior survey administered annually by IPA.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

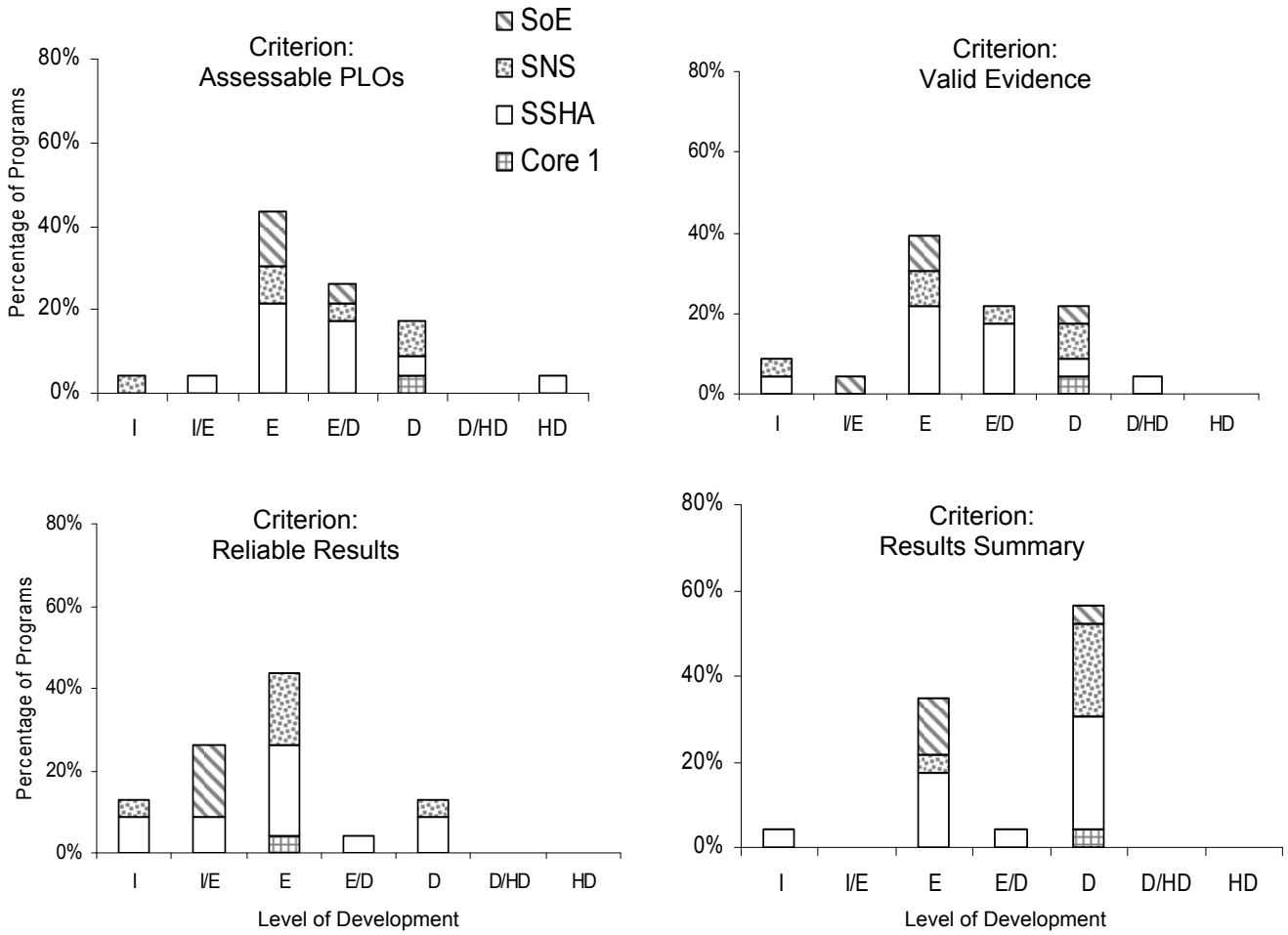
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



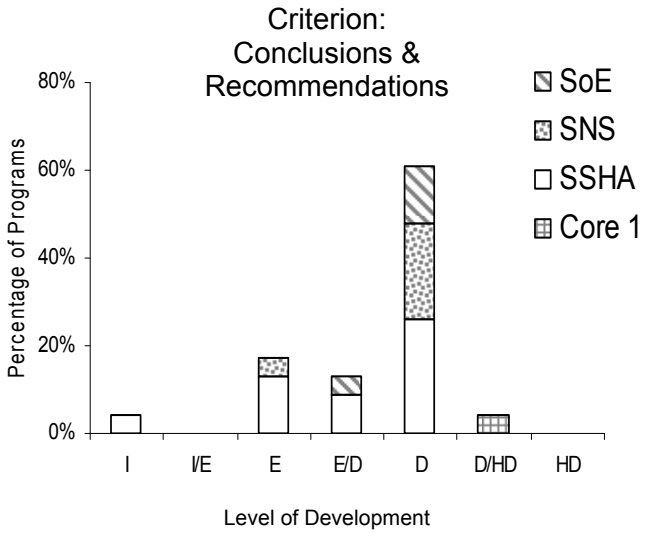


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

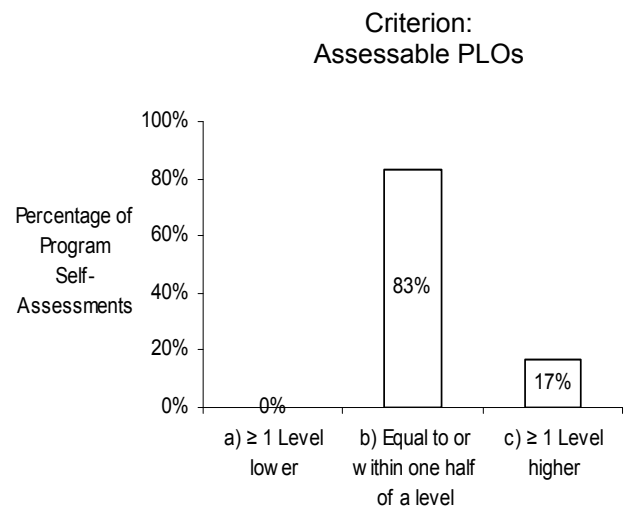
These results indicate that

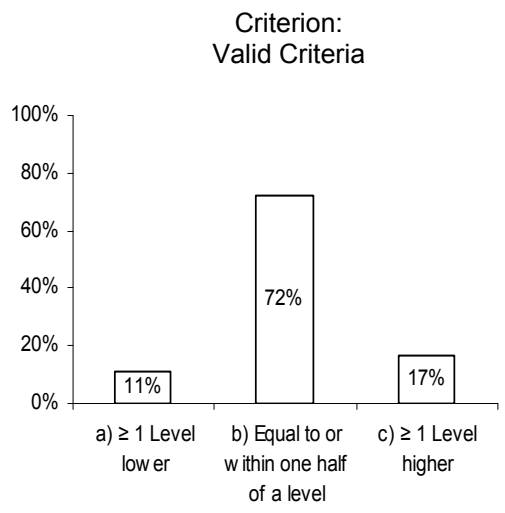
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

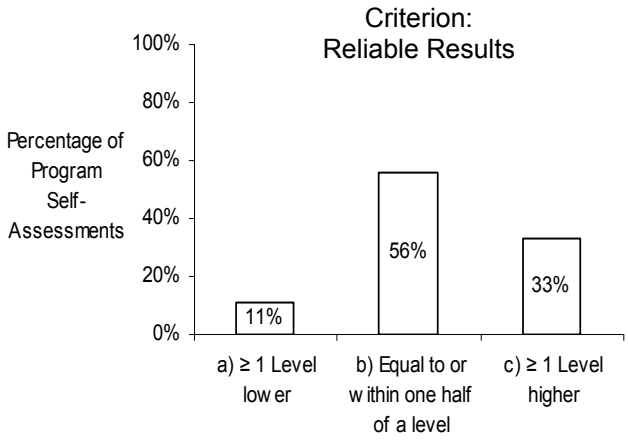
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

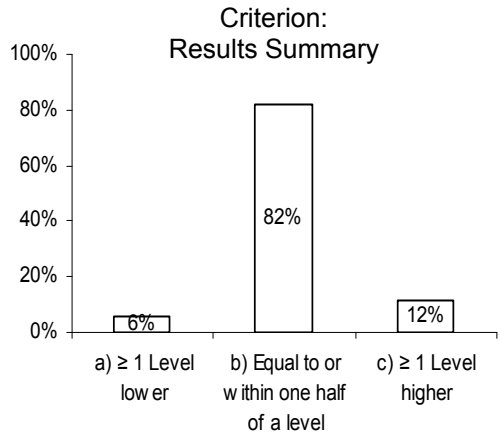
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









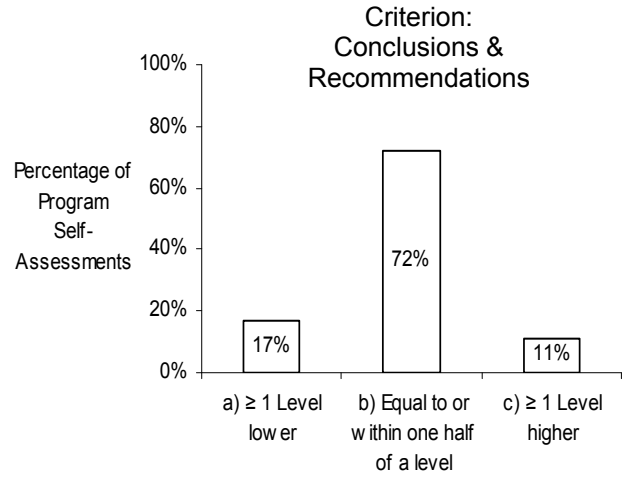


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?

UNIVERSITY OF CALIFORNIA

BERKELEY • DAVIS • IRVINE • LOS ANGELES • MERCED • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

UNIVERSITY OF CALIFORNIA, MERCED
P.O. BOX 2039
MERCED, CALIFORNIA 95344

2 September 2010

Dear Virginia:

Please find attached feedback regarding your program's 2009 PLO Assessment Report, provided by the Ad-Hoc Committee for Review of PLO Assessment Reports, as well as the committee's final report to the WASC Steering Committee.

Convened in support of UC Merced's Educational Effectiveness Review, the ad-hoc committed was asked to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment,* 3) identify common assessment or student learning-related strengths, weaknesses or issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

Using a structured review format, the committee identified the primary strengths of each program's reported assessment work and a couple of key areas for further development. The committee hopes that this feedback will help to strengthen your program's assessment practices as we continue to implement assessment as a tool for improving our students' learning.

The committee was also keen to direct your program to relevant assessment-related resources. Since the most productive resource would be individualized support, the committee awaits a decision on recommendations for resourcing assessment that have been made to the Provost by the Senate-Administrative Council on Assessment (SACA). In the interim, the committee notes that several PLO Reports highlighted the SATAL program (Students Assessing Teaching and Learning) as a useful way to gain indirect evidence of student learning and feedback on how well the program is meeting student needs. Finally, this fall WASC is offering a Student Learning and Assessment Retreat, which may be of use to your program. The Provost is able to support registration and travel costs.

On behalf of the Steering Committee, I thank you very much for all your hard work in support of our Initial Accreditation effort. We hope that this feedback helps to advance your program's assessment efforts.

Sincerely,

Gregg Camfield
Professor and Chair, WASC Steering Committee

PLO Report Review Summary

Name of Program: **SPANISH MINOR**

1) Reviewers' assessments of the program's level of development for each of the five criteria in the *Rubric for the Report on PLO Assessment*.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Average	**HD**	**D/HD**	**E/D**	**D**	**D**
Primary	**HD**	**D/HD** – have only engaged in assessment once but clearly committed to meaningful revisions and improvements	**E/D** – faculty are familiar with rubrics and regularly use them but are they calibrating regularly? they check for agreement but aren't necessarily independently assessing	**D/HD** – benchmarking will evolve within the program over years but other results are clearly summarized	**D**
Secondary	**HD –** a useful rubric and an assessable PLO, although the latter could be streamlined and could also be foregrounded better in the report	**D/HD** – excellent dedication to the intricacies of student written expression; sample size might be larger, but holistic assessment of student work makes up for it	**E/D –** there's a clear commitment to and shared understanding of the rubric, although it's not clear how faculty were calibrated, and explanation of student work tended to be general	**D –** careful holistic attention to student work, with qualitative summary (though there might be more quantitative information, to specify how students measured up)	**D –** thoughtful examination of assessment method, although could be elaborated or projected more specifically with respect to future efforts

2) A clear strength of the program's assessment efforts.

Primary Reviewer: A clear strength of the program's assessment effort is the alignment of the program's design, implementation, and assessment with ACTFL proficiency guidelines, which provide external validation of the program's student learning achievements.

Secondary Reviewer: The program's holistic, shared approach to student learning and program assessment is admirable.

3) Two or three assessment practices to be strengthened, based on the criterion (or criteria) identified in question 1 as needing the most development.

Primary Reviewer: (1) This is difficult. The work is assessment work is exemplary virtually without exception. The consistency of the results suggests that that the unavoidably small sample size is likely to be less of an issue in having confidence in the results. The proposed improvements flow logically from the reported work and results. As the program itself suggests, it would be of value to involve more faculty in the assessment. This would not only reduce the workload but would also directly increase faculty familiarity with student strengths and weaknesses, providing an opportunity to consider this first hand knowledge in the implementation of their own courses. (2) It would be helpful to have a sense of how often raters disagreed and if there were areas in which disagreements were more likely to occur. This would help to establish program standards for calibration, which will serve the program as it grows.

Secondary Reviewer: Given the predominantly positive results (i.e., primarily "excellent" or "good"), I sometimes wondered about the extent to which they may be inflated (and the assessment exercise not be as rigorous as it might be). Clearer calibration of raters could help in this regard, as can developing familiarity with using the rubric, and more specific explanation in student portfolio comments (which tended to be somewhat general). To this end, the report might further specify how student work measures up to the rubric.

Final Report of the
Ad-hoc Committee for Review of 2009 Academic Program Learning Outcome Assessment Reports
September 2, 2010
Prepared by Laura E. Martin
WASC Coordinator

Introduction

This report summarizes the work and results of the ad-hoc committee to review the 2009 Academic Program Learning Outcome Assessment Reports (PLO Reports) formed under the auspices of the WASC Steering Committee in Spring 2010, and charged to 1) provide formative feedback to individual academic programs regarding their first PLO assessment efforts, 2) establish baseline data describing the degree to which academic assessment efforts are developed as evaluated using the *Rubric for the Report on PLO Assessment* (Appendix A), 3) identify common assessment or student learning-related strengths, weaknesses or potential issues as potential foci for further study or action, and 4) identify program assessment efforts to be highlighted in the EER Report.

The committee reviewed twenty-three PLO Reports, representing 15 undergraduate majors, six stand alone minors, one graduate program, and Core 1 (Appendix B). Reports for four undergraduate majors and 2 stand alone minors[1] were unavailable at the time of review. The committee's conclusions, results, and methods follow.

Conclusions

As represented in the *Rubric for the Report on PLO Assessment,* a foundational purpose of annual assessment and reporting is to cultivate an intentional, transparent, collaborative, and thereby programmatic approach to fostering student intellectual development within a degree granting program. Over time, these activities should result in a well-aligned and integrated set of courses (curriculum) taught to support student achievement of a comprehensive set of program learning outcomes (PLOs), for which the criteria and standards of student performance are detailed more specifically through instruments like programmatic rubrics.

The *Rubric for the Report on PLO Assessment* supports development of these attributes through criteria and standards that ask faculty to

- Collaboratively develop a shared set of explicit expectations for student intellectual development by graduation (*Assessable PLO* and *Valid Evidence* criteria).
- Collaboratively develop and apply tools to investigate student development and achievement of these expectations that should be a product of the program's curriculum rather than a single course (*Valid Evidence* and *Reliable Results* criteria).
- Take action as a program to improve student intellectual achievements and to refine the assessment practices (*Results Summary* and *Conclusions and Recommendations* criteria).

[1] Majors: Bioengineering, Economics, Management, and Materials Science and Engineering. Minors: American Studies and Service Science. These latter programs have fewer than 5 students each.

- Record these actions and decisions to allow current and future faculty to connect their own course curricula and instructional practices to these results as appropriate (all criteria).

Programs should be working to reach the *Developed* standard.

Based on the aggregated results of report reviews, the ad-hoc committee came to the following conclusions related to the degree of development of assessment practices across academic programs and associated needs.

1) Two notable strengths of the reported assessment practices were the widespread use of assessment results to inform curricular change and, for many programs, the degree of faculty involvement in programmatic assessment activities.

2) For most programs, continued development is needed to achieve a transparent, collaborative, programmatic approach to assessment. Development efforts should be focused primarily on the practices associated with three foundational criteria of the rubric - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*.

3) A majority of programs need to develop one or more lines of indirect evidence of student learning that complements the direct evidence gathered.

4) To support development, programs should continue to evaluate their annual assessment practices by using the *Rubric for the Report on PLO Assessment.*

5) To support development, efforts should be made to further establish a common assessment-related vocabulary and set of expectations and tools across academic programs.

Related to student learning, the committee recognized that

6) Communication skills, particularly writing, may benefit from development efforts focused at the disciplinary level. Further investigation is warranted.

Additionally, the committee noted that it will be important to

7) Update the annual PLO Assessment Report format to ensure that programs archive with their report the rubric used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO as well as the examination of student learning achievements during program review. WASC will also expect this material to be available in future accreditation reviews.

8) Remind programs that have revised PLOs to update catalog, websites, and syllabi and consider the implications of these modifications for the alignment of the program's curriculum across courses and for course-level student learning outcomes.

Finally, the committee identified the following programs as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

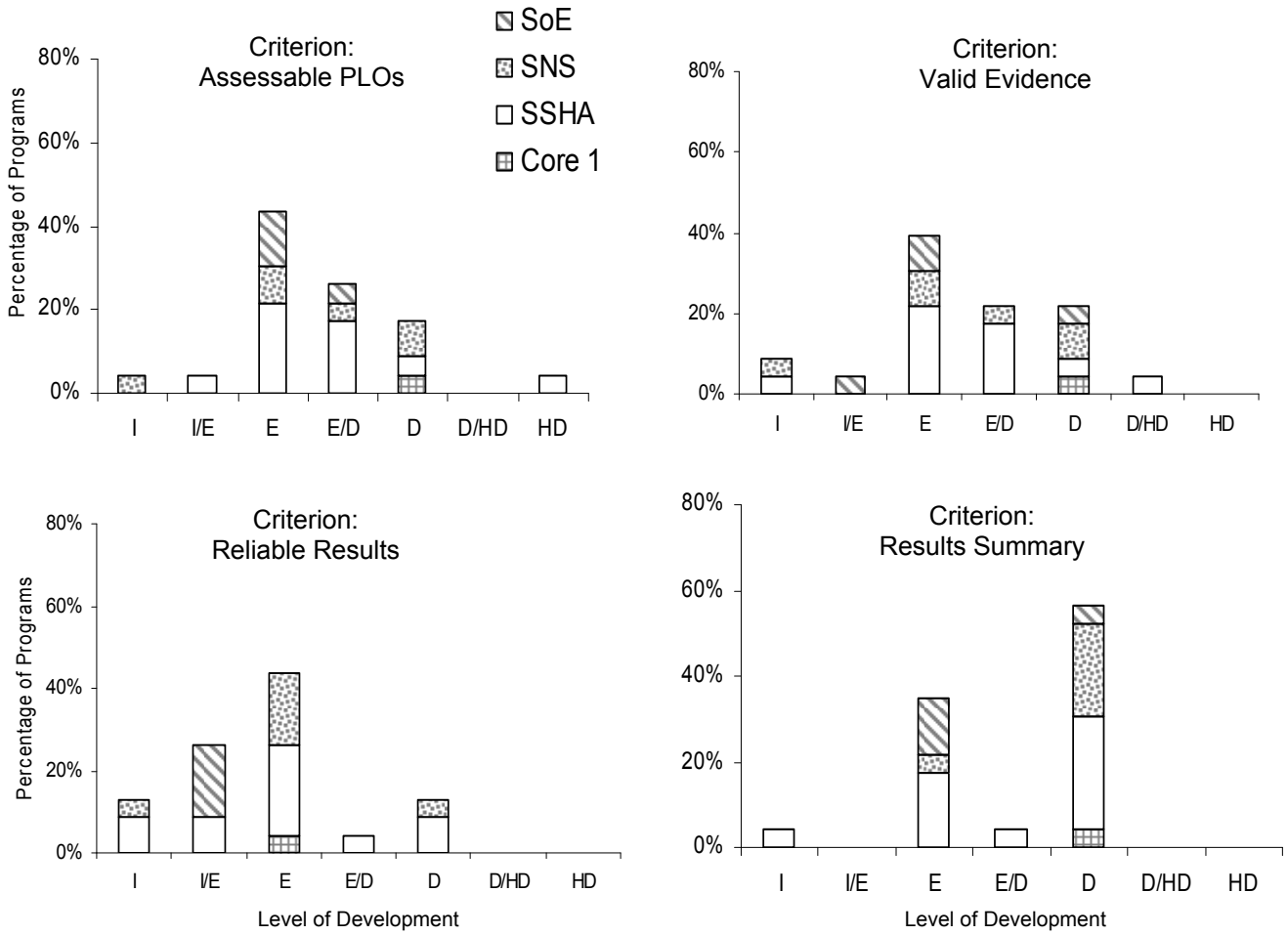
Specific details related to be the above conclusions are available the Results section of this report.

Results

Establishing Baseline Data on Program Assessment

PLO Reports communicated a wide diversity of assessment-related strengths, many of which were unique to individual programs. Broad faculty involvement was the most frequently recognized strength; it was noted in response to 39% of the reports.

Figure 1 summarizes the degree of program development for each of the rubric's five criteria, in the aggregate and by School, as assessed by the ad-hoc committee.



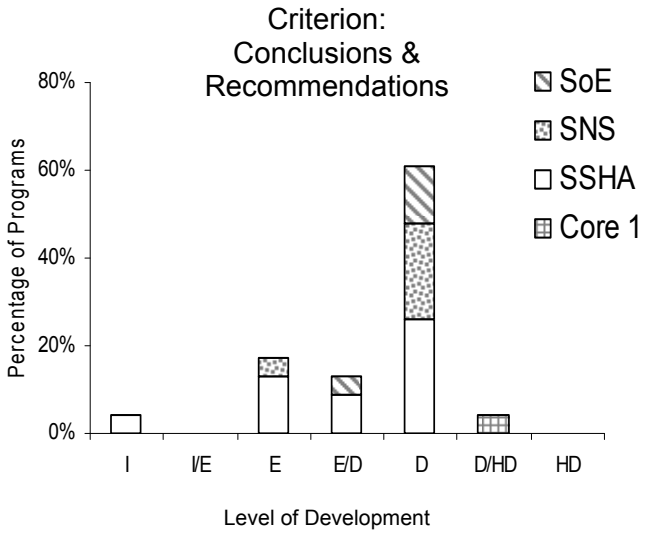


Figure 1. Results of the committee-based assessment of the level of development of the 2009-2010 Program PLO Reports as evaluated using UC Merced's *Rubric for the Report on PLO Assessment*. For each of the rubric's five criteria, the percentage of reports judged to be at each level of development - Initial (I), Emerging (E), Developed (D), Highly Developed (HD) or intermediate to two levels (for example, I/E) - is summarized below by School (n=23).

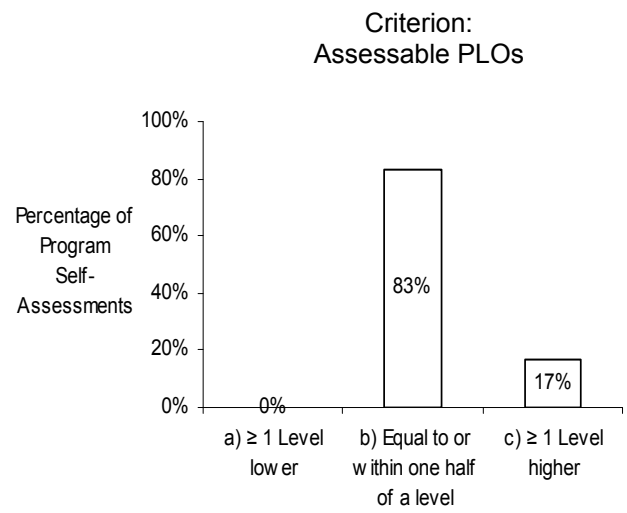
These results indicate that

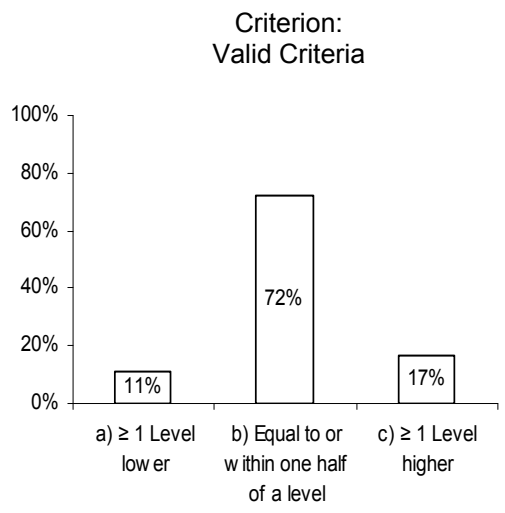
1) In the aggregate, UC Merced's academic programs are moving toward *Developed* assessment structures and practices, with most reporting programs assessed to be *Emerging* or better for all five of the rubric's criteria.

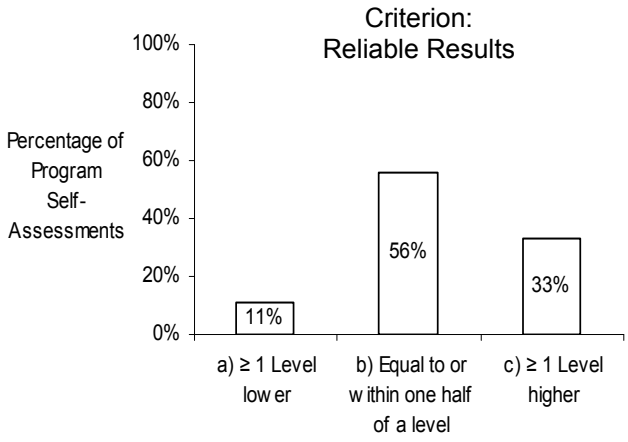
2) The criteria associated with the key foundational practices of transparent, collaborative programmatic practices - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results* - are most in need of development, with approximately 50% of programs assessed as *Emerging* or lower in these categories.

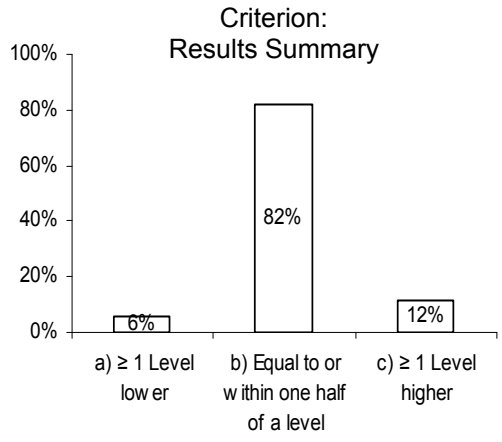
3) Assessment results are being connected to curriculum and instructional practices even in these earliest stages, as a majority of programs (~60%) reached the *Developed* or *Highly Developed* standard with respect to the *Results Summary* and *Conclusions and Recommendations* criteria. Indeed, approximately 90% of reporting programs identified changes to be made to the curriculum as a result of the assessment process.

Figure 2 summarizes the degree of agreement between the committee's assessment of program development and each program's self-assessment with respect to the rubric's five criteria.









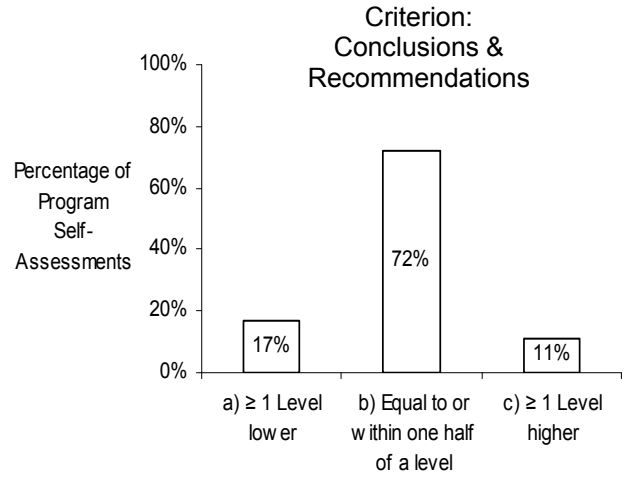


Figure 2. Comparison of program and committee assessments of PLO Reports using UC Merced's *Rubric for the Report on PLO Assessments*. For each of the rubric's five criteria, a figure below describes the percentage of program self-assessments were a) at least one level of development lower than the committee's assessment, b) the same as or with-in one-half level of committee's assessment, and c) at least one level of development higher than the committee's (n=18 for all criteria, except Results Summary with an n=17).

The strong agreement between committee and program assessments for nearly all criteria

1) Lends support to the conclusions regarding the current level of assessment-related development among reporting academic programs.

2) Indicates that the rubric is supporting the development of a shared, foundational, understanding of the expectations of assessment practices and reporting across academic programs.

The agreement between committee and program assessments was weakest for the *Reliable Results* criterion, with a third of program self-ratings exceeding those of the committee by at least one level of development. Allowing for the possibility that low committee scores and relatively high program self-ratings may reflect abbreviated reporting rather than actual practice, the results suggest that many programs need to formalize the calibration (norming) process. This will increase confidence in the conclusions of assessment work and promote shared understandings that feedback into course-based instruction to better support achievement of programmatic curricular goals.

To more precisely identify the assessment practices in need of development, Table 1 provides the most common suggestions made to programs and their relationship to the three criteria specifically identified for attention - *Assessable PLOs*, *Valid Evidence*, and *Reliable Results*. Suggestions 1, 2, 5, and 6 describe practices essential to establishing a collaborative, programmatic approach to improving student learning through assessment.

Table 1: The most common suggestions for improving assessment practices in relation to the rubric criteria each supports.

Committee Suggestion	**Rubric Criteria Addressed by Committee Suggestion**		
	Assessable PLOs	**Valid Evidence**	**Reliable Results**
1) Further elaborate rubrics to describe more precisely the criteria and standards used to evaluate student work so that a) faculty engaged in the assessment process are using shared rather than personal criteria to evaluate student work, and b) faculty, including future faculty, who are not directly involved in the assessment process can see their colleagues' expectations for student learning in relation to a PLO.	x	x	
2) In the absence of a capstone course or assignment, strategically select examples of student work from a number of courses so that the assessment results enable conclusions about the program's efficacy in bringing about student learning as opposed to that of a single course. Alternatively, explain why student work from a single course is expected to represent learning facilitated by the program's curriculum, rather than just that of the course.		x	

Committee Suggestion	Rubric Criteria Addressed by Committee Suggestion		
	Assessable PLOs	Valid Evidence	Reliable Results
3) Consider more explicitly the alignment between the assessed student work and the PLO. Do the assignments elicit student responses that allow the program to evaluate achievement of the PLO?	x	x	
4) Develop lines of indirect evidence to complement and support results of direct assessment. [2]		x	
5) Formalize calibration processes in order to increase confidence in conclusions and develop shared understandings to take back to individual courses.			x
6) Determine inter-rater reliability in order to, as necessary, take steps to improve agreement among faculty and confidence in conclusions.			x

Emerging themes

The committee identified several recurrent issues or themes related to both student learning and assessment.

With respect to student learning, several programs noted that

- Student written communication skills and mathematical expression confounded the assessment of student knowledge and intellectual skills.

With respect to assessment, the committee noted the need to

- Develop a programmatic approach to assessment that involves identifying student work that reflects learning due to the program's curriculum rather than that of a single course.
- Develop explicit rubrics that are appropriate/well-aligned to the assessment task.
- Continue to address the challenge of assessing forms of evidence that are not easily susceptible to criteria, for example, complex narrative.
- Develop indirect evidence of student learning that complements direct evidence.
- Improve faculty calibration.

[2] Only 43%, of the 95% of programs that worked with direct evidence, used some form of indirect evidence as well. The absence of indirect evidence contributed in part to approximately 60% of programs being identified as emerging or emerging/developed for the *Valid Evidence* criterion. Programs that gathered indirect evidence, particularly through student interviews, typically described these results as being very informative and as practices that will be continued. In several cases, student perspectives on the program independently supported faculty observations or conclusions based on direct evidence of student performance. WASC also expect multiple lines of evidence.

Programs to Potentially Highlight in the Educational Effectiveness Review Report

Based on the quality of the assessment work and results, the following programs were identified as potential candidates to be highlighted in the Educational Effectiveness Review Report: Core 1, Environmental Engineering, History, Mechanical Engineering, Physics, and Spanish Minor. The committee recognized that the final selection would depend upon the larger context of the report.

Methods

The ad-hoc committee included six individuals, consisting of faculty and staff from each School and Core 1/General Education (Appendix C). Each PLO Report was reviewed by a primary and secondary reviewer guided by a template (Appendix D) constructed around the *Rubric for the Report on PLO Assessment* (Appendix A). Following an in-person norming session focused on an example report, reviews were conducted asynchronously with primary reviewers forwarding completed reviews to secondary reviewers. After reading the PLO Report, secondary reviewers then considered the primary reviewer's summary comments and, as necessary, noted any supplemental or discrepant points. Primary and secondary review responsibilities were split evenly among committee members. Workload was distributed in this way to reduce the workload associated with reading and evaluating a large number of reports near the end of the semester.

Through this process, reviewers

1) Rated the program's level of development for each of the rubric's five criteria;
2) Identified two or three assessment practices to be strengthened, based on the rubric criteria identified in step 1 to be most in need of development; and
3) Identified one to two strengths of each program's assessment practices.

For the few cases in which the two reviewers' rubric-based ratings disagreed by one level of development (ex. emerging versus developed), the average rating was calculated (ex. emerging/developed). For half-step differences in rater scores (ex. intermediate/developed versus developed), the shared level of development was calculated (ex. developed). Reviewers never disagreed by more than one level of development.

To identify frequently observed assessment or student learning-related strengths, weaknesses or potential issues, reviewers' narrative comments were coded and the frequency of each code was calculated. Using these results, the committee identified common assessment or student learning-related themes or issues to be addressed during an in-person meeting.

Finally, to gain some sense of how useful the rubric is for communicating assessment-related expectations and to gauge each program's impression of the quality of its assessment work, the committee's rubric scores were compared to the rubric-based self-evaluations each program reported in its PLO Report.

Appendix A: UC Merced's *Rubric for the Report on PLO Assessment*.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

	Criterion	Initial	Emerging	Developed	Highly Developed
ASSESSMENT METHODS	**Assessable Program Learning Outcome (PLO)**	PLO does not identify what students can do to demonstrate learning (vague, immeasurable verb statements like "students understand major theories"). No rubric developed.	PLO indicates how students can demonstrate learning. Action verb may be general and the PLO may not be observable or measurable. Assessment criteria[1] have not been identified or are incomplete. Rubric in early stages of development.	PLO describes how students can demonstrate learning, identifying observable and measurable results. Criteria are articulated in the form of a rubric, criteria and standards[1] may need further development to be more meaningful and consistently applied.	PLO specifically describes how students can demonstrate learning. Rubric clearly articulates explicit criteria and standards[1] for assessing the PLO, identifies the most important aspects of student learning, and includes descriptions of student performance at varying levels.
	Valid Evidence	It is not clear that potentially valid evidence is collected for the PLO <u>and/or</u> individual faculty use personalized rather than programmatic criteria and standards[1] to assess student work or performance.	Faculty have reached general agreement on the types of evidence to be collected for the PLO but may not include both direct and indirect forms. Evidence needs to be further focused or aligned with PLO or emerging criteria to produce truly meaningful and useful results.	Faculty collect relevant & sufficient evidence for each outcome, including both indirect and direct evidence. Assessment instruments (ex. rubric) assess the level of student attainment. Evidence is aligned with the PLO and assessment criteria to enable meaningful results and conclusions.	Assessment criteria have been pilot-tested and refined over time, usually shared with students. Direct and indirect evidence are designed to mutually inform conclusions. Feedback has led to refinements in the assessment process.
	Reliable Results	Reviewers of student work are not calibrated to apply assessment criteria in a uniform way; there are no checks for inter-rater reliability	Reviewers are calibrated to apply assessment criteria in a uniform way <u>or</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way <u>and</u> faculty routinely check for inter-rater reliability.	Reviewers are calibrated, and faculty routinely find assessment data to have high inter-rater reliability.
RESULTS & CONCLUSIONS	**Results Summary**	Results (data table or other means) are not included in report.	Results (data table or other means) are included but unclear or missing key data.	Results clearly delineated for each line of evidence in tabular or other summary formats. May reference benchmarks or other expectations.	Results clearly delineate each line of evidence, indicating various levels of achievement. Includes benchmarks.
	Conclusions & Recommendations	Report identifies implications but no recommendations for improvement in student learning or assessment practices and no explanation of how these claims are derived. No reasoning offered in support of claims.	Report identifies some conclusions, implications, and recommendations for improvements regarding student learning or assessment, but the claims are vague or questionably related to results. Support for claims is occasionally insufficient. Questions of validity or reliability are not discussed. Results may be discussed by limited number of faculty, recommendations may be difficult to implement due to lack of convincing results and/or limited faculty involvement or support.	Report clearly articulates conclusions, implications and recommendations for improvement regarding both student learning and assessment and which could be drawn from results. Includes some consideration of the reliability and validity of results. May offer vague support for some claims. Results have been discussed by many faculty and recommendations likely to be implemented due to faculty involvement and support and quality of assessment work.	Report articulates a well-reasoned critique of conclusions, implications, and recommendations that could be drawn from the results for both student learning and assessment. Includes a well-reasoned discussion of validity and reliability of results. Faculty discuss results, plan needed changes, secure necessary resources, and implement changes. Efforts to collaborate with others, such as librarians or student affairs professionals, to improve results.

[1] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Laura E. Martin & Anne Zanzucchi, Center for Research on Teaching Excellence, University of California, Merced. 2009.

RUBRIC FOR REPORT ON PLO ASSESSMENT

Center for Research on Teaching Excellence

This rubric has five major criteria:

(1) **Assessable Program Learning Outcomes**: Program learning outcome should be reasonable and appropriate for the degree level. If national disciplinary standards are available, the PLO may reflect those priorities. To be assessable, it should involve specific, active verbs with supporting details describing how students will demonstrate their learning. For instance, avoid verbs of general cognition such as "know" or "understand" and instead use verbs like "demonstrate by" or "solve" that show how learning is applied. Through discussion of examples of student work and perhaps course-specific rubrics used by faculty, faculty groups have agreed on explicit criteria[2] and elaborated a program-level rubric. For more information, see <http://crte.ucmerced.edu/program-learning-outcomes-resources>.

(2) **Valid Evidence**: To be valid, evidence must be discussed among faculty and aligned with both the expectation(s) described by the PLO and the criteria[2] faculty use to evaluate student learning. Valid evidence is also linked to sample size or sampling approach, so as to be representative of a norm. For more information, see the appended selection on sample sizing from Linda Suskie's *Assessing Student Learning: A Common Sense Guide* (2004).

(3) **Reliable Results**: Reliable results stem from agreement among faculty about the standards[2] used to evaluate student work, usually as articulated in a faculty-developed, program-wide rubric. Agreement about how to apply these standards in the evaluation of student work (i.e. calibration) is rooted in discussion and practice. Some questions to consider are: How do faculty promote calibration? How do faculty check for calibration? I.e. when faculty apply a rubric to student work, how consistently do they reach the same conclusions (i.e. exhibit inter-rater reliability)? If results are inconsistent, how can inter-rater reliability be improved?

(4) **Summarizing Results**: When drafting a results chart (data table or other means), it is important to consider multiple audiences. How would faculty within your department understand the results? If viewed by outside stake-holders like students, faculty from other programs, administrators, parents, etc., would they reach similar conclusions? Comparing the results to previous results in your program, expectations your program has set for student learning, or to results of similar programs within or outside of the UC (i.e. benchmarking) can provide context for interpreting the results.

(5) **Conclusions and Recommendations**: An effective conclusion closes the loop by analyzing results and implementing change. The narrative should address some probable conclusions based on the results. For example, if students were not given a clear incentive to participate in a particular assessment, the results may not be completely reliable as students may not have been motivated to perform at their best. Specific actions and a timeline for implementation should also be provided since the goal is to gather data to improve both student learning and the ability to engage in effective assessment. Changes might include improving the assessment process or curriculum, examining curriculum content in support of skill development, changing pedagogical practices, stimulating faculty discussion, simply re-examining program learning outcomes, or identifying ways student support services (tutoring, advising, the library) might contribute to increased student success.

---

[2] Criteria are the specific skills or abilities to be measured. Standards describe the levels of performance for a given criterion (ex. proficient to exemplary) and in doing so enable their measurement.

Based on C. Jenefsky & JFKU Program Review Council (2008) and WASC rubrics (2007)

Appendix B: 2009 PLO Report Submission Record as of Spring 2010 when the review was conducted.

School	Program Name	Program Type	2009 PLO Report Submitted?
SoE	Bioengineering	Major	No
	Computer Science & Engineering	Major	Yes
	Environmental Engineering	Major	Yes
	Materials Science & Engineering	Major	Yes
	Mechanical Engineering	Major	No
	Environmental Systems	Graduate	Yes
SNS	Applied Mathematics	Major	Yes
	Biology	Major	Yes
	Chemistry	Major	Yes
	Earth Systems Science	Major	Yes
	Physics	Major	Yes
	Natural Sciences Education Minor	Minor	Yes
SSHA	Anthropology	Major	Yes
	Cognitive Science	Major	Yes
	Economics	Major	No
	History	Major	Yes
	Literatures & Cultures	Major	Yes
	Management	Major	No
	Political Science	Major	Yes
	Psychology	Major	Yes
	Sociology	Major	Yes
	American Studies	Minor	Yes
	Arts	Minor	Yes
	Philosophy	Minor	Yes
	Spanish	Minor	Yes
	Service Science	Minor	Yes
	Writing Minor	Minor	Yes
College One	Core 1	GE	Yes

Appendix C: Roster of the Ad-hoc Committee for PLO Report Review, Spring 2010.

Name	Title & Relevant Committee Memberships	School/ Unit Affiliation
Gregg Camfield	Professor, Chair of the WASC Steering Committee; member Senate-Administrative Council on Assessment	SSHA
Tom Hothem	Lecturing Faculty; member of the WASC Steering Committee	SSHA & Core 1 (GE)
Valerie Leppert	Associate Professor	SoE
Laura Martin	WASC Coordinator, Assessment Coordinator, member of the WASC Steering Committee, member Senate-Administrative Council on Assessment	Center for Research on Teaching Excellence
Masa Watanabe	Director of Student Success, School of Natural Sciences	SNS
Jack Vevea	Associate Professor, member of the WASC Steering Committee	SSHA

Appendix D: Review Template

PLO Report Review: Instructions and Form

Background:
The goals of this PLO Assessment Report Review are to (1) provide feedback to programs on their assessment efforts, (2) identify and report back to each School's faculty any assessment or student learning issues common to the School's programs, and (3) identify programs whose results might serve as case studies in our EER Report. To support this work, we will also (4) rate each program's assessment efforts against the *Rubric for the Report on PLO Assessment*.

Instructions:
Primary Reviewers:
For each PLO Report you review please complete the primary reviewer sections of the Review Form, then forward the completed forms to the secondary reviewer.

Secondary reviewers:
Please review the PLO Reports and the primary reviewer's responses to the Review Form. In the secondary reviewer sections of the form, please note any differences with the primary reviewer's conclusions, or any additional thoughts, you might have. Finally please submit completed Review Forms to Laura, lmartin@ucmerced.edu.

PLO Report Review Form

1) Name of Program:_____

2) Please assess the program's level of development with respect to each of the five criteria in the *Rubric for the Report on PLO Assessment* (Appendix B). Provide your conclusions, along with any supporting comments, in the table below as I (Initial), E (Emerging), D (Developed) or HD (Highly Developed). A program can be assessed to fall between two levels of development, for example, I/E or E/D.

Reviewer	Assessable PLO	Valid Evidence	Reliable Results	Results Summary	Conclusions & Recommendations
Primary					
Secondary					

3) Please provide the program with constructive feedback regarding its assessment practices. (These comments will be excerpted and shared with the program on behalf of this committee, so please craft these with your colleagues in mind.)

   a) In one sentence, describe a clear strength of the program's assessment efforts.

      Primary Reviewer:

      Secondary Reviewer:

   b) Based on the criterion (or criteria) identified in question 2 as needing the most development, and the corresponding supplemental questions provided in Appendix A, please identify two or three assessment practices to be strengthened.

      Primary Reviewer:

      Secondary Reviewer:

4) Please note what you might imagine to be emerging, shared themes related to the assessment process and/or student learning results.

Primary Reviewer:

Secondary Reviewer:

5) Briefly evaluate the potential of the program's assessment work as a case study for our EER report. Relevant criteria include:
   - Quality of assessment work, including most importantly evidence of assessment-based actions to improve student learning.
   - Illustrative of a commonly observed approach to assessment.
   - Illustrative of trends or common conclusions emerging from PLO Reports.

Primary Reviewer:

Secondary Reviewer:

6) Any outstanding thoughts or questions?

Primary Reviewer:

Secondary Reviewer:

APPENDIX A: A set of questions is provided below to help guide the identification of assessment practices to be strengthened in response to Question 3 above. To support this process, the questions are organized by the criteria that appear on *Rubric for the Report on PLO Assessment*. *Please pay particular attention to italicized questions as they were specifically identified in the most recent WASC Commission Action Letter as important areas of development and so should have priority in feedback.*

Assessable PLOs:

- *As written, is the PLO measurable? Does it involve specific, active verbs that such as "demonstrate by" or "solve" as opposed to verbs of general cognition like "know" or "understand"?*
- Is the PLO likely to be understood by students? Of use to students?
- To help faculty (and students as relevant) develop a shared understanding of what student mastery of the PLO looks like in practice, has a rubric been developed that articulates criteria[3] and standards[4] of performance (for each criterion)?

Valid Evidence:

- Is a rationale for the assessment strategy provided? Does the program explain why a particular piece of work, or a particular course, is an appropriate focus for examining student achievement with respect to the PLO?
- Related to the bullet above, does the assessment work have a program/PLO focus rather than course-level focus?
- *Does the assessment method include at least one form of direct evidence (i.e. actual student work)?*
- Is the assessment measure going to produce results that bear on the PLO? (I.e. Is it aligned with the PLO?)
- Will the sample size and sampling strategy produce results that represent the student norm?
- Are multiple, complementary forms of evidence used to more precisely identify areas in need of attention and to strengthen confidence in the conclusions? (For example, direct and indirect evidence?)

Reliable Results:

- Did the program use a rubric with explicit standards and criteria to review student work and, thereby, promote agreement among reviewers about student proficiency?
- Did at least two faculty members review each piece of student work?
- Were faculty reviewers calibrated or normed with respect to explicit standards and criteria used to asses student work in order to promote agreement among reviewers about observed student proficiencies?
- Did the program determine how consistently faculty reached the same conclusion with respect to a piece of student work (i.e. determine inter-rater reliability)?

Summarizing Results:

[3] "The qualities we look for in student evidence." (Driscoll and Wood, 2007) The specific skills or abilities to be measured.
[4] Standards describe the levels of performance for a given criterion (ex. proficient to exemplary).

- To gain a sense of the distribution of student performance relative to performance standards or levels of proficiency, does the program describe the percentage of students meeting specific levels of performance, for example, as described in a rubric?
- Does the program identify a goal for the percentage of students meeting minimum or higher levels of proficiency? Are the assessment results evaluated in relation to this goal?

Conclusions and Recommendations:

- Are the program's conclusions supported by the results?
- Are issues related to the validity and reliability of the results considered in drawing conclusions and identifying actions to be taken on the basis of those conclusions?
- As warranted, does the program propose some actions to be taken in response to their conclusions? Are the actions well-aligned with the conclusions?
- In order to promote improvements in student learning have the results, conclusions and proposed actions been shared with the faculty and approved by the faculty?