Report Format for PLO Assessment

Format Recommendations

Due: January 30, 2010

Context: For the Educational Effectiveness Review (EER), the second stage of the application for Initial Accreditation, UC Merced must evaluate from an institutional perspective 1) how effectively assessment is practiced across the institution and, 2) what has been revealed through assessment about our students' learning outcomes relative to our expectations. The purpose of this process is to show change and identify improvements. Your PLO Assessment Report is essential to this effort and will be submitted as evidence in our EER Report, represented as aggregate data.

Purpose: This guide will help you in preparing the report **due January 30, 2010** that describes your program's assessment of its first program learning outcome. Much like a professional journal article, *the content of the report should be focused by the assessment question the program pursued*. *As such, your report will only address a subset of the questions provided below: those that reflect the purpose(s) of the assessment work conducted.*

Audience: Although accreditation is the impetus for initiating annual assessment, these PLO reports (and those submitted annually hereafter) will provide a history of your program's planned evolution with respect to improving student learning as well as the program's ability to engage in effective, meaningful assessment. As such, please write these reports with your present and future colleagues in mind, aiming to document what your program has learned both about student learning and about the process of engaging in assessment in order to promote continued improvement in both of these areas. Your PLO report will also highlight any revision of your program's curriculum and specific evidence of student learning that prompted a change.

Submission: Specific information will follow.

Support: *If you have questions about any aspect of these guidelines*, please contact the member of the Center for Research on Teaching Excellence who has been assigned to support your program. (See appended list.) Also appended is the useful resource "Rubric for the Report on PLO Assessment". Page length guidelines are provided for each section. The report, excluding appendices, should total approximately eight to ten double-spaced pages in length.

I. Abstract (1 paragraph)

Briefly summarizes the question the program pursued, the methods, results, conclusions, recommendations and resource implications.

II. Introduction (1 page)

Introduces readers to the program's goal(s), with discussion as to how and why the Program Learning Outcome (PLO) was selected as the first to review. What assessment question did the program pursue? Specifically, what did your faculty want to know and achieve?

III. Assessment Methods (2 pages)

In general, what methods were employed to gather and evaluate evidence/data to address the assessment question? This section describes the assessment methods and process used to evaluate student learning with respect to the PLO, including the process for evidence collection and analysis. Supporting details that provide insight into the validity and reliability of the results should be described. These might include: (1) sampling size and strategy (ex. population sampled and incentives for participation), (2) pilot-tested and revised rubrics (ie.

rubrics that have been used and revised to refine descriptions and improve usefulness), (3) faculty calibration (i.e. faculty collaboratively develop rubrics with explicit criteria and standards and employ methods for promoting inter-rater reliability, or high-levels of agreement, in scoring), etc. How and to whom results and conclusions were disseminated as well as how next steps were decided should be included here. Relevant dates and meeting times are important to include.

IV. Results (2 pages)

This section summarizes in written and tabular and/or graphical form the results of the analyses of the direct and indirect evidence of student learning with respect to the PLO. As appropriate and meaningful, quantitative summaries should be included as should performance expectations against which these results will be evaluated. The following are some ways to approach describing results: (1) Are your students meeting your program's performance expectations, for example, as described by a rubric? What percentages at what levels of performance? (2) How do your students compare to peers (majors to non-majors, to other UC majors, or to students at comparable non-UC institutions)? (3) Are your students improving? How many and how so? (4) Are students doing as well as they can? If not, what is their potential capability?

V. Conclusions & Recommendations (2-3 pages)

A. *Student Learning*: Describes the implications of the results, including integration of direct and indirect evidence of student learning and an analysis of the effect assessment methods may have on the validity and reliability of the results. Recommends actions to improve student learning with respect to desired intellectual skills and knowledge as well as a timeline for implementation. Actions may fall into one or more of the following categories or perhaps others: (1) instruction, (2) curriculum, (3) course sequencing, (4) co-curricular support for student learning (ex. tutoring), or (5) communicating expectations to students.

B. *Assessment Methods*: What worked and did not work in this assessment process? What can be improved immediately and what has been established as long-term goals? What advice would you provide to your colleagues with respect to this assessment strategy for future use? Consider issues like the precision of the research question, appropriateness of the evidence, factors affecting the 'assessability' of the PLO and validity of the results. Related to the latter two items: is the verb of the PLO sufficiently precise to promote shared performance expectations among faculty and students and meaningful assessment? 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. Do faculty share a common understanding of what a particular level of performance looks like, i.e. are the faculty "calibrated" so that different individuals will reliably draw the same conclusions about the quality of student work?

VI. Implications of Proposed Changes (Planning / Budget) (1 page)

Are there any resources that will be needed to implement the above plans? How and where will be the resources obtained?

VII. Self Evaluation (1 page)

Using the appended "Rubric for the Report on PLO Assessment", evaluate your program's level of development for each criterion (far left column). Please include a brief explanation for each rating.

VIII. Appendices

Please include relevant supporting documents, including *as appropriate* rubrics, brief meeting summaries, curriculum, summary reports, memos, and sample student work.

Appendix: FAO Staff Support Listings

School of Engineering

Program	FAO	CRTE & IPA Support Staff
Bioengineering	Ariel Escobar	Laura Martin & Nancy Ochsner
Environmental Engineering	Tom Harmon	Laura Martin & Nancy Ochsner
Computer Science & Engineering	Ming-Hsuan Yang	Karen Dunn-Haley & Mike Roona
Materials Science & Engineering	Christopher Viney	Laura Martin & Nancy Ochsner
Mechanical Engineering	Carlos Coimbra	Anne Zanzucchi & Nancy Ochsner

School of Natural Sciences

Program	FAO	CRTE & IPA Support Staff
Applied Mathematics	Arnold Kim	Anne Zanzucchi & Nancy Ochsner
Biological Sciences	Patti LiWang	Laura Martin & Nancy Ochsner
Chemical Sciences	Anne Kelley	Anne Zanzucchi & Nancy Ochsner
Earth Systems Science	Lara Kueppers	Laura Martin & Nancy Ochsner
Physics	Kevin Mitchell	Anne Zanzucchi & Nancy Ochsner
NSED Minor	Arnold Kim	Anne Zanzucchi & Nancy Ochsner

School of Social Sciences, Humanities and Arts

Program	FAO	CRTE & IPA Support Staff
Anthropology	Kathleen Hull	Karen Dunn-Haley & Mike Roona
Political Science	Tom Hansford	Mike Truong & Mike Roona
Psychology	Will Shadish	Karen Dunn-Haley & Mike Roona
Cognitive Science	Jeff Yoshimi	Mike Truong & Mike Roona
Economics	Shawn Kantor	Pending

Program	FAO	CRTE & IPA Support Staff
History	Sholeh Quinn	Karen Dunn-Haley & Mike Roona
Literatures & Cultures	Gregg Camfield	Mike Truong & Mike Roona
Management	Hans Bjornsson	Robert Ochnser & Nancy Ochsner
American Studies Minor	Gregg Camfield	Mike Truong & Mike Roona
Arts Minor • Media Arts • GASP	Dunya Ramicova Kevin Fellezs	Karen Dunn-Haley & Mike Roona
Philosophy Minor	Peter Vandershraaf	Mike Truong & Mike Roona
Sociology Minor	Irenee Beattie	Karen Dunn-Haley & Mike Roona
Writing Minor	Anne Zanzucchi	Anne Zanzucchi & Nancy Ochsner
Service Science Minor	Hans Bjornsson	Robert Ochnser & Nancy Ochsner
Spanish Minor	Virginia Adan-Lifante	Adriana Signorini & Mike Roona
Core 1	Tom Hothem	Anne Zanzucchi & Nancy Ochsner
Core 100	Anne Zanzucchi & Tom Hothem	Anne Zanzucchi & Nancy Ochsner

Graduate Programs

Program	FAO	CRTE & IPA Support Staff
Environmental Systems	Tom Harmon	Laura Martin & Nancy Ochsner

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 and/or 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 or faculty routinely check for inter-rater reliability.	Reviewers are calibrated to apply assessment criteria in a uniform way and 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)

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)

Assessing Student Learning
A Common Sense Guide

Linda Suskie
Towson University

Anker Publishing Company, Inc.
Bolton, Massachusetts

Some published test "batteries" for basic (K–12) education are normed so that achievement scores can be compared against aptitude scores, but these kinds of tests are rare at the higher education level. This perspective is therefore seldom used in higher education, except informally.

If a capability perspective is appropriate for your situation, design your assessment to collect information systematically on student capability as well as achievements. Then compare your students' achievements against their capability, and decide if their achievement levels are satisfactory. This perspective usually requires subjective, professional judgment rather than statistical analysis to decide whether student performance levels are acceptable.

Which Perspective Should You Use?

Each perspective—standards-based, benchmarking, best practice, value-added, longitudinal, and capability—has advantages and disadvantages and each gives a somewhat incomplete picture of student learning. *Using multiple perspectives—as many as appropriate and feasible—will give you the most complete portrait* of how effectively your students are achieving your goals. If you are examining the performance of your students on a certification examination, for example, you might look at how they compare against national norms, against students in the best programs in the country, and against your students who took the examination a few years ago. Considering all these perspectives will give you a more balanced picture of how your students are performing.

How Much Evidence Should You Examine?

Obviously the more assessment evidence you collect and consider, the greater confidence you will have in your conclusions about student learning. Faculty who look at 300 essays will have more confidence in their conclusions about student writing skills than faculty who look at ten essays. But more evidence means more precious time spent collecting and examining it, so an important question is, "How much evidence is enough?"

To get a perspective on how much evidence you should examine when assessing program or institutional goals, consider that professional pollsters, trying to determine the opinions of hundreds of

millions of Americans, rarely survey more than 1,000 people. If you've seen the results of such surveys, you may have noticed that pollsters note an "error margin" of about 3%. This means that if a pollster finds, for example, that 76% of Americans think the President is doing a good job, the pollster is very sure (actually 95% sure) that, if all Americans could be surveyed, between 73% and 79% (76% plus and minus 3%) would say that the President is doing a good job. Table 6.6 lists the error margins of various sample sizes.

Table 6.6

Error Margins of Various Sample Sizes

Random Sample Size	Error Margin
9,604	1%
2,401	2%
1,067	3%
600	4%
384	5%
264	6%
196	7%

While professional pollsters often aim for samples of about 1,000 people, with an error margin of 3%, unless your assessments may lead to major (read expensive) changes at your institution, a sample of 300 or 400 is probably sufficient.

What if 300 or 400 is still too large to be practical? What if faculty only have the time to score, say, 100 essays? Yes, you can use smaller sample sizes—whatever number you think is feasible—if you recognize that your sampling error will be larger and if you take careful steps to ensure the accuracy and truthfulness of your assessment findings (see Chapter 2).

What if you have a small program or institution? Table 6.7 lists the sample sizes you'd need for 5% sampling errors from some relatively small groups of students.

Table 6.7

Sample Sizes Needed From Small Groups for 5% Error Margins

Number of Students You're Sampling From	Random Sample Size
1,000	278
500	217
350	184
200	132
100	80
50	44

How Might a Sample Be Chosen?

Simple random samples—in which every student has an equal chance of being selected—are a straightforward way to ensure that your sample is representative of all your students. A simple random sample might be drawn by writing every student's name on a separate slip of paper, putting all the slips in a bag, shaking the bag, and drawing out as many names as you need. This can be done virtually by using software to generate a random sample. If such software isn't available (check with your computer center), select students based on the last few digits of their student identification numbers, as these last few digits are usually randomly distributed. If you have 250 students and wish to examine writing samples from 50 (20%) of them, for example, you could choose all students whose student identification numbers end in, say, 4 or 5 (20% of all possible digits 0 through 9).

Simple random samples aren't always practical. If you want to administer an in-class survey, for example, it wouldn't be feasible to choose a random sample of the entire student body, go to every class, and ask just the sampled students in each class to complete the survey while the rest of the students sit idle. If a simple random sample isn't realistic, other kinds of samples are possible.

Cluster random samples involve taking a random sample of subgroups of students and then collecting information from everyone in those subgroups. You could take a random sample of first-year writing classes, for example, and then assess essays written by everyone in those classes. Or you could take a random sample of floors in the residence halls and interview everyone on those floors.

Purposeful or judgment samples are carefully *but not randomly* chosen so that, in your best judgment, they are representative of the students you are assessing. Suppose that you want to assess essays written by students in first-year writing classes and you would like to select a cluster random sample of classes. Unfortunately, you know that while some faculty will gladly cooperate, others will decline to provide you with copies of student essays. You can still obtain a good sample of essays by choosing, from those classes with cooperating faculty, a sample of classes that meet on various days and at various times and seem to represent a good cross-section of all first-year writing classes.

Or suppose that you want to assess student learning in general education science courses and, while students may meet this requirement by taking any of 17 courses, 80% take one of just two courses: introductory biology and introductory geology. Collecting assessment information from just these two courses will be far simpler and more cost effective than collecting information from all 17 courses and will give you useful information on 80% of your students.

If you must use a small or non-random sample or if you have a low participation rate, it's especially important to collect information showing that your sample is representative of students in general (see Chapter 14).

Report Format for PLO Assessment

Format Recommendations

Due: January 31, 2011

Purpose: This guide will help you to prepare the report **due January 31, 2011** that describes your program's assessment of its second program learning outcome. Much like a professional journal article, *the content of the report should be focused by the assessment question the program pursued*. *As such, your report will only address a subset of the questions provided below: those that reflect the purpose(s) of the assessment work conducted.*

Audience: PLO reports are intended to provide a record of your program's planned evolution with respect to improving student learning as well as the program's ability to engage in effective, meaningful assessment. As such, please write these reports with your present and future colleagues in mind, aiming to document what your program has learned both about student learning and about the process of engaging in assessment in order to promote continued improvement in both of these areas. Your PLO report will also highlight any revision of your program's curriculum and specific evidence of student learning that prompted a change.

***Submission*: Please submit your report to your Dean no later than January 31, 2011.**

I. Abstract (1 paragraph maximum)

Briefly summarize the question the program pursued, the methods, results, conclusions, recommendations and resource implications.

II. Introduction (1 page maximum)

Introduce readers to the program's goal(s). Briefly discuss how and why the Program Learning Outcome (PLO) was selected for review. What assessment question did the program pursue? Specifically, what did your faculty want to know and achieve?

III. Assessment Methods (2 pages maximum)

Describe the assessment methods and process used to evaluate student learning with respect to the PLO, including the process for evidence collection and analysis. In general, what methods were employed to gather and evaluate evidence/data to address the assessment question?

Supporting details that provide insight into the validity and reliability of the results should be described. These will likely include: (1) the number of lines of evidence and their forms[1], (2) sample size and sampling strategy[2], (3) pilot-tested and revised rubrics[3], and (4) methods for faculty calibration[4]. How and to whom results and conclusions were disseminated as well as how next steps were decided should be included here. Relevant dates and meeting times are useful to include.

---

[1] Does the study involve direct and indirect sources of evidence regarding student achievement of the PLO? Does the evidence allow the program to examine the impact of the *program's* curriculum on student learning?

[2] What are the demographic characteristics of the students sampled? For example, are they mostly seniors? Were the students motivated to produce high quality work that reflects their abilities? For example, were the assessments required coursework (i.e. embedded) or were students asked to volunteer for the assessment?

[3] Rubrics that have been used and revised to refine descriptions and improve usefulness.

[4] Faculty calibration, also called norming, involves collaboratively developing rubrics with explicit criteria and standards and, in the review of student work, using methods to promote inter-rater reliability (high-levels of agreement among individuals in the scoring of student work).

IV. Results (2 pages maximum)

Summarize in written and tabular (or graphical) form the results of the analyses of the direct and indirect evidence of student learning with respect to the PLO. As appropriate and meaningful, quantitative summaries should be included as should performance expectations against which these results will be evaluated. The following are some ways to approach describing results: (1) Are your students meeting your program's performance expectations, for example, as described by a rubric? What percentages at what levels of performance? (2) How do your students compare to peers (majors to non-majors, to other UC majors, or to students at comparable non-UC institutions)? (3) Are your students improving? How many and how so? (4) Are students doing as well as they can? If not, what is their potential capability?

V. Conclusions & Recommendations (2-3 pages maximum)

A. *Student Learning*: Describe the implications of the results, including the integration of direct and indirect evidence of student learning and an analysis of the effect assessment methods may have on the validity and reliability of the results. Recommend actions to improve student learning with respect to desired intellectual skills and knowledge as well as a timeline for implementation. Actions may fall into one or more of the following categories or perhaps others: (1) instruction, (2) curriculum, (3) course sequencing, (4) co-curricular support for student learning (ex. tutoring, library instruction, etc.), or (5) communicating expectations to students.

B. *Assessment Methods*: Describe what worked and did not work in this assessment process and how it can be improved. Identify practices that can be improved immediately and those to be established as long-term goals. Consider issues like the precision of the research question, appropriateness of the evidence, factors affecting the 'assessability' of the PLO, and validity of the results. In relation to the latter two items, consider the verb of the PLO. Is it sufficiently precise to promote shared performance expectations among faculty and students and meaningful assessment? Active verbs like "demonstrate by…" or "solve," that show how learning is applied, support student learning (and its assessment) more effectively than verbs of general cognition such as "know" or "understand". Additionally, do faculty share a common understanding of what a particular level of performance looks like, i.e. are the faculty "calibrated" so that different individuals will reliably draw the same conclusions about the quality of student work?

VI. Implications of Proposed Changes (Planning / Budget) (1 page maximum)

Are there any resources that will be needed to implement the above plans for improvement? How and where will be the resources obtained?

VII. Self Evaluation (1 page maximum)

Using the appended "Rubric for the Report on PLO Assessment", evaluate your program's level of development for each criterion (far left column). Please include a brief explanation for each rating.

VIII. Appendices

Please be sure to append any rubrics used to evaluate student work together with representative examples of scored student work. This will support calibration and comparison when revisiting the PLO in the future as well as the examination of student learning during program review. WASC will also expect this material to be available in future accreditation reviews. Similarly, your program might also consider including *as appropriate* relevant meeting summaries, summary reports, or memos.

RUBRIC FOR THE 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[5] 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.

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

RUBRIC FOR THE 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[6] 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.

---

[6] 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.