Initial Assessment Project Summary - Reassessment Report
Form Only

LAC Focal Assessment Projects are intended to result in improved student learning, not to document how successful they already are. As a result, our projects follow a multi-stage cycle of "assess - correct - reassess." Briefly, but very specifically provide the essential information from the already completed "assess - correct" stages of this ongoing assessment project here.

<table>
<thead>
<tr>
<th>Initial Assessment Project Summary (previously completed assessment project)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Briefly summarize the main findings of your initial assessment. Include either 1) the frequencies (counts) of students who attained your benchmarks and those who did not, or 2) the percentage of students who attained your benchmark(s) and the size of the sample you measured:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Briefly summarize the changes to instruction, assignments, texts, lectures, etc. that you have made to address your initial findings:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>If you initially assessed students in courses, which courses did you assess:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>If you made changes to your assessment tools or processes for this reassessment, briefly describe those changes here:</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Section 1: Outcome for Focal Analysis

1. Outcome Chosen for Focal Analysis

1A. Briefly describe what and why this focal outcome is being investigate: (e.g., "First term students do not seem to be able to transfer the knowledge from their math class to our program class. We wish to investigate student understanding of the needed math concepts upon entry into our course. If students do have the theoretical understanding, we will investigate ways we can help students apply their knowledge in a concrete application." A second example is: "Anecdotally, it seems that our first year students are not retaining critical information between Winter and Spring Quarters." We will measure student benchmark attainment in Winter Quarter.

1B. If the assessment project relates to any of the following, check all that apply:

- Degree/Certificate Outcome – if yes, include here: 
- PCC Core Outcome – if yes, which one: 
- Course Outcome – if yes, which one: 
1. It is essential to have a clear statement of how the assessment project relates to improving student attainment of a core outcome. Keep this statement brief and specific. Strive for clarity - this element is central to the assessment project.

2. Enter all degree/certificate outcomes that are assessed in some way in this project.

3. Enter as many PCC Core Outcomes as have been mapped to the pertinent degree/certificate outcomes. A list of the Core Outcomes can be found at http://www.pcc.edu/resources/academic/core-outcomes/.

4. Enter any course outcomes that are being assessed in some way in this project.

Section 2: Project Description

1. Most faculty assessment projects involve assessing student performance in one or more courses. Include any course(s) involved in the current assessment.

2. Include the number and type of instructors who will likely be teaching this course at the time the assessment is conducted. Having adequate representation of all instructors is essential to good assessment.

3. Include all the course outcomes from the CCOG(s) for each course listed above that are relevant to your measurement of this core outcome. CCOGs can be found at: http://www.pcc.edu/ccog/. In many cases, SACs can use assignments that are already being given and graded/evaluated in their courses to assess core outcomes (these sorts of assignments are called 'embedded assessments'). For this approach to work effectively, some course outcomes must align with some core outcomes.

4. Use this section to indicate the type of assignment/assessment. Check all that apply. Whenever possible/feasible, attach the activity description/instructions/etc. as an appendix to the Annual Plan. For some help with selecting and creating optimal assessments, see: http://assessment.aas.duke.edu/documents/DirectandIndirectAssessmentMethods.pdf
Section 2: Project Description - Type of Measurement

2B. How will you score/measure/quantify student performance?

[ ] Rubric (used when student performance is on a continuum - if available, attach as an appendix – if in development - attach to the completed report that is submitted in June)

[ ] Checklist (used when presence/absence rather than quality is being evaluated - if available, attach as an appendix – if in development - attach to the completed report that is submitted in June)

[ ] Trend Analysis (often used to understand the ways in which students are, and are not, meeting expectations; trend analysis can complement rubrics and checklist)

[ ] Objective Scoring (e.g., Scantron scored examinations)

[ ] Other – briefly describe: [ ]

2C. Type of assessment (select one per column)

[ ] Quantitative

[ ] Qualitative

[ ] Direct Assessment

[ ] Indirect Assessment

If you selected ‘Indirect Assessment’, please share your rationale: [ ]

Qualitative Measures: projects that analyze in-depth, non-numerical data via observer impression rather than via quantitative analysis. Generally, qualitative measures are used in exploratory, pilot projects rather than in the assessments of student attainment. Indirect assessments (e.g., surveys, focus groups, etc.) do not use measures of direct student work output. These types of assessments are also not able to truly document student attainment.
1. Typically in educational assessment, student performance is evaluated and somehow scored/tracked. Select all that apply to this project.

**Rubrics** require that one or more dimensions/trait be measured (often called "criterion") in a scaled way (often called "level"). PCC has identified four general levels for most of our core outcomes. If you click on one of the core outcomes after directing your web browser to [http://www.pcc.edu/resources/academic/core-outcomes/](http://www.pcc.edu/resources/academic/core-outcomes/) you will in most cases see the four levels defined and illustrated for that core outcome. The illustrations are only examples: each SAC may interpret and apply the core outcome level in the most appropriate way for their discipline and particular circumstances. Faculty often desire that their students demonstrate the highest level of outcome attainment: level four. PCC has not established a policy for this and SACs are responsible for identifying the appropriate desired and required levels of student outcome attainment. As a beginning point of discussion, an analysis of our core outcome levels in the context of a more national expectation for community college student outcome attainment (as elaborated in the [Degree Qualifications Profile](http://www.aacu.org/toolkit/)) suggests that our level 3 is a likely realistic expectation for outcome attainment in our students.

A well-designed and carefully normed rubric (one that leads to useful, consistent scores) is essential to many core outcome assessment projects. The following links point to helpful information about creating rubrics:

- [http://leap.aacu.org/toolkit/](http://leap.aacu.org/toolkit/) (the AACU's Leap Campus Tool-kit)
- [http://www.learningoutcomesassessment.org/documents/DQPop1.pdf](http://www.learningoutcomesassessment.org/documents/DQPop1.pdf) (Details an interesting approach to outcomes assessment)

If you have specific questions or need additional help, contact your LAC assessment coach.

**Checklists** can be used in assessment when the purpose is to track the presence/absence (dichotomous scoring) of student attainment, rather than when scaled scoring is performed (based on multiple possible levels of student performance). For example, did hand-washing occur both before and after patient contact? Checklists can sometimes be easier to create and to score, though care should be taken in their creation and use. (a useful checklist for checklists: [http://www.ecu.edu/cs-educ/opd/upload/KeystoQAChecklist.pdf](http://www.ecu.edu/cs-educ/opd/upload/KeystoQAChecklist.pdf))

**Trend Analysis**: Rubrics can lose important information about the student work. For example, after scoring student work, it may not be apparent if all students who were scored at “level 2” on a rubric have similar issues. A Trend Analysis can give a more nuanced understanding of the student work and the common errors students are making (and the strengths students have).

How to do a Trend Analysis:

- Evaluate a portion of the student work to find the common “trends.” These could be things students are missing or things they are doing well. Make a list of “trends.”
• Decide which trends are most important to consider.
• Faculty members count the number of occurrences for each trend item. If there are a lot of trends, it is helpful to have faculty members focusing on different trends.
• If work has been scored by a rubric, it could be helpful to see if particular rubric scores have the same trends (i.e., see if students who scored a “2” have similar mistakes) or if student make similar mistakes for all rubric levels.

Other commonly used measures include:


**Interviews/Surveys** are almost always indirect measures of student outcome achievement. As a result, they tend to be less useful measures. Given both the practical difficulties of directly measuring some core outcomes (due to both the nature of the outcome and the costs/difficulties associated with direct assessment), they can be acceptably measured with carefully constructed, indirect surveys or interviews. For some tips on survey construction, see: [http://help.surveymonkey.com/articles/en_US/kb/Design-Tips-How-to-create-and-administer-effective-surveys](http://help.surveymonkey.com/articles/en_US/kb/Design-Tips-How-to-create-and-administer-effective-surveys). The LAC has a Qualtrix account that faculty members can use to conduct complex surveys - though we currently have minimal ability to support SAC’s use of this new tool. Contact your LAC coach for more information. For surveys not requiring complex statistical analysis, Survey Monkey personal accounts may be useful.

### 2. Quantitative and Qualitative Measures

**Quantitative Data:** involves numbers and is often used to generalize results from a sample to the targeted population of interest. **Qualitative Data:** is non-quantitative and is used to gain an understanding of the reasons or motivations that might provide insight into a particular problem. This type of analysis can be very time consuming, so only a small number of non-representative cases are used. It is often used to explore complex inter-relationships between variables: often with the intention of identifying new questions for further research. The analysis is non-statistical and can rarely be used to make reliable generalizations about the population of interest. Though, it can be used very effectively in exploratory or orienting research, as a starting point in a comprehensive research/assessment program, etc. Examples include: in-depth interviews, group discussion, focus groups, etc.

Combining quantitative and qualitative data gives rich and appropriately actionable knowledge to address barriers in student success.

**Example:** A college collects data from the student population focusing on student success rates. A random sample of 700 students is used to ensure a representative sample of the various ethnic groups on the campus (quantitative data) is obtained. Results of the study show that Group X has a lower success rates than other groups. The college decides to do a focus group of 20 students from Group X to understand the group’s perceived barriers to success (qualitative data).

### 3. Direct and Indirect Assessment
**Direct Assessment:** A direct assessment involves looking at samples of student work and evaluating the work to determine if the student has shown competency for a learning outcome or skill. Direct assessments include the following:

- Evaluation of a test or essay designed to demonstrate student knowledge, understanding, application, analysis, or criticism
- Capstone project
- Watching a student perform a skill such as welding a pipe or listening to a pulse to detect a murmur

**Indirect Assessment:** An indirect assessment is a measure of a perception of attainment of a competency/outcome. The perception can be the student’s or another individual's. Indirect assessments include the following:

- Asking a class of students to give a “thumb’s up” if they understand the concept/skill
- Survey
- Focus group
- Exit interviews
- Grading homework for completion (students may have worked every problem wrong)
- Attendance (presence in class does not necessarily mean students have gained understanding nor does their absence necessarily mean that they do not have the understanding)

**Example:** At a medical exam, we typically fill out a form about our family history of various diseases and the doctor usually asks us how we are feeling. Sometimes we are not honest with these indirect assessments. But even if we are honest, they are, at best, a guide to help the doctor diagnosis our aliment. For instance, a patient may think he has stomach cancer because his symptoms are like his beloved aunt (or are listed under stomach cancer on some internet website), but he really has a bad ulcer. The direct assessments are the various tests the doctor may order (urine, stress test, colonoscopy, biopsy) that will hopefully lead to pinpointing the actual problem. We can’t “lie” about the direct assessment, but this doesn’t mean it always gives us helpful information (the right test needs to be performed and the test needs to be accurate).

**Interviews/Surveys** are almost always indirect measures of student outcome achievement. As a result, they tend to be less useful measures. Given both the practical difficulties of directly measuring some core outcomes (due to both the nature of the outcome and the costs/difficulties associated with direct assessment), they can be acceptably measured with carefully constructed, indirect surveys or interviews. For some tips on survey construction, see: [http://help surveymonkey com/ articles/en_US/ kb/Design-Tips-How-to-create-and-administer-effective-surveys](http://help surveymonkey com/ articles/en_US/ kb/Design-Tips-How-to-create-and-administer-effective-surveys). The LAC has a Qualtrix account that faculty members can use to conduct complex surveys - though we currently have minimal ability to support SAC’s use of this new tool. Contact your LAC coach for more information. For surveys not requiring complex statistical analysis, Survey Monkey personal accounts may be useful.

4. Most experts recommend using direct assessments - whenever possible - when assessing for student attainment of outcomes. However, there are times when indirect assessment is desirable or much more feasible. Share your rationale for using an indirect assessment here.
Sample Rubric

1. Sample rubric. 3 levels are often sufficient to properly evaluate student achievement. Notice that this rubric score of 3 - Competent - corresponds to PCC's Core Outcomes Level 3. In this example, competence is the SAC expectation for acceptable introductory college student performance.

| Artifacts or work samples that do not satisfy the criteria for Level 2 Developing should be scored 1. |
|--------------------------------------------------|--------------------------------------------------|--------------------------------------------------|
| **Student identifies and uses sample characteristics to evaluate the conclusions of a scientific study.** | **Excellent 4** | **Competent 3** | **Developing 2** |
| Sample characteristics are interpreted and evaluated accurately in developing a comprehensive analysis or synthesis. Authors' conclusions are thoroughly questioned. | Sample characteristics are interpreted and evaluated accurately in developing a coherent analysis or synthesis. Authors' conclusions are subject to questioning. | Sample characteristics are addressed but are not coherently analyzed or effectively included in a synthesis. Authors' conclusions are accepted or dismissed without adequate explanation. |
| **Student identifies and uses design characteristics to evaluate the conclusions of a scientific study.** | Design characteristics are interpreted and evaluated accurately in developing a comprehensive analysis or synthesis. Authors' conclusions are thoroughly questioned. | Design characteristics are interpreted and evaluated accurately in developing a coherent analysis or synthesis. Authors' conclusions are subject to questioning. | Design characteristics are addressed but are not coherently analyzed or effectively included in a synthesis. Authors' conclusions are accepted or dismissed without adequate explanation. |
| Student includes and accurately interprets p-values, confidence intervals, and effect sizes to evaluate the conclusions of a scientific study. | Identified quantitative elements are interpreted and evaluated accurately in developing a comprehensive analysis or synthesis. Authors' conclusions are thoroughly questioned. | Identified quantitative elements are interpreted and evaluated accurately in developing a coherent analysis or synthesis. Authors' conclusions are subject to questioning. | Identified quantitative elements are addressed but are not coherently analyzed or effectively included in a synthesis. Authors' conclusions are accepted or dismissed without adequate explanation. |

Section 2: Validity Indicators

2D. Check any of the following that were used by your SAC to create or select the assessment/scoring criteria/instruments used in this project:

- Committee or subcommittee of the SAC collaborated in its creation
- Standardized assessment
- Collaboration with external stakeholders (e.g., advisory board, transfer institution/program)
- Theoretical Model (e.g., Bloom’s Taxonomy)
- Aligned the assessment with standards from a professional body (for example, The American Psychological Association Undergraduate Guidelines, etc.)
- Aligned the benchmark with the Associate’s Degree level expectations of the Degree Qualifications Profile
- Aligned the benchmark to within-discipline post-requisite course(s)
- Aligned the benchmark to out-of-discipline post-requisite course(s)
- Other (briefly explain: __________)

An assessment is valid when it accurately assesses what it is intended to. Select all that apply to your SAC’s assessment tools.
Section 2: Timing and Purpose

1. Indicate the timing of your planned assessment here.
2. This group is usually called the 'population.' Each SAC will determine who their population is for each project. For example, one SAC might be interested in generalizing their results to all students who complete their introductory sequence. Another SAC might be interested in all students who complete a service learning project in any of their courses. It helps to be explicit about the population of interest: generally, good samples cannot be picked until the population is clearly identified.
3. When your SAC is measuring a clearly established outcome and/or intend to use the results of an assessment to drive programmatic change, measuring the population (when small, e.g., a graduating cohort in a degree program) or a representative sample is essential. Mark the first box in section H and proceed to the next section.

If your SAC is conducting a preliminary or exploratory investigation (usually the first step in a more comprehensive assessment program - one intended to test the waters or to develop tentative hypotheses to further explore later (often quantitatively) - it might make sense to select a very small (and as a result unrepresentative sample) for in-depth, qualitative investigation. If your SAC is using this strategy, share your rationale for selecting your sample here.
Section 2: Sampling

21. Which will you measure?

1. the population (all relevant students – e.g., all students enrolled in all currently offered sections of the course)
2. a sample (a subset of students)

If you are using a sample, select all of the following that describe your sample/sampling strategy (refer to the Help Guide for assistance):

1. Random Sample (student work selected completely randomly from all relevant students)
2. Systematic Sample (student work selected through an arbitrary pattern, e.g., “start at student #1 on the roster and then select every 5th student following”, repeating this in all relevant course sections)
3. Stratified Sample (more complex, consult with an LAC coach if you need assistance)
4. Cluster Sample (students are selected randomly from meaningful, naturally occurring groupings (e.g., SES, placement exam scores, etc.))
5. Voluntary Response Sample (students submit their work/responses through voluntary submission, e.g., via a survey)
6. Opportunity/Convenience Sample (only some of the relevant instructors are participating)

The last three options in bolded red have a high risk of introducing bias. If your SAC is using one or more of these sample/sampling strategies, please share your rationale:

22. Briefly describe the procedure you will use to select your sample (including a description of the procedures used to ensure student and instructor anonymity. For example:

We chose to use a random sample. We asked our administrative assistant to assist us in this process and she was willing. All instructors teaching course XXX will turn-in all student work to her by the 9th week of Winter Quarter. She will check that instructor and student identifying information has been removed. Our SAC decided we wanted to see our students' over-all performance with the rubric criteria. Our administrative assistant will code the work for each section so that the scored work can be returned to the instructors (but only she will know which sections belong to which instructor). Once all this is done, I will number the submitted work (e.g., 1-300) and use a random number generator to select 56 samples (which is the sample size given by the Kaiso sample size calculator for 300 pieces of student work). After the work is scored, the administrative assistant will return the student work to individual faculty members. After this, we will set up a face-to-face meeting for all of the SAC to discuss the aggregated results.”

We chose to use a random sample. We asked our administrative assistant to assist us in this process and she was willing. All instructors teaching course XXX will turn-in all student work to her by the 9th week of Winter Quarter. She will check that instructor and student identifying information has been removed. Our SAC decided we wanted to see our students’ over-all performance with the rubric criteria. Our administrative assistant will code the work for each section so that the scored work can be returned to the instructors (but only she will know which sections belong to which instructor). Once all this is done, I will number the submitted work (e.g., 1-300) and use a random number generator to select 56 samples (which is the sample size given by the Kaiso sample size calculator for 300 pieces of student work). After the work is scored, the administrative assistant will return the student work to individual faculty members. After this, we will set up a face-to-face meeting for all of the SAC to discuss the aggregated results.”
1. The group that you want to generalize to is known technically as the 'population.' The population is determined by the purposes of the assessment. In some cases, the population is the '34 students scheduled to complete our program at the end of spring term.' In others, it might be 'all students who will take the introductory course in our discipline.' If you are measuring your population of interest, select the appropriate check box and continue directly to Section 3. If you are measuring a subset of your population of interest, select the 'sample' check box and continue filling-out this section.

2. Before explaining the specifics of this section, some common background needs to be established. When we sample, we are attempting to measure a subset of the total population in order to accurately generalize to it. Normally, sampling is done when the population is too large to measure in any reasonable way. To make the complex story about sampling short, the real-world limitations of educational assessment at PCC make true representative sampling 'extremely difficult' to 'practically impossible' (e.g., student characteristics can change in significant ways with changing economic conditions, the academic term, recent events, etc. - careful conclusions from data gathered this term or year may not apply to students the next term or year). Further, all of the technical recommendations for conducting research (e.g., sample size suggestions, appropriate p-values, etc.) are based on the assumption that a true random sample of the population has been made. Since this can rarely happen in our assessment projects, determining the best sampling practices for our work is extremely difficult.

For these reasons, we are adopting a recommended approach to sampling that should maximize the accuracy and utility of our real-world projects while not over-burdening instructional faculty with artificially rigorous standards. To begin, the 'sample or sampling frame' is the actual group from which our sample is drawn. For example, if our population of interest is 'all 201a students' the sample frame is the 'complete list of students enrolled in all sections of the course when we pull our sample'. Reflection on the nature of the population and the sample frame reveals that in most cases, the two are not identical. (Note: we have immediately violated the assumptions underlying all of the standard recommendations surrounding good research.) Problematically, these (and other) violations of the underlying assumptions make the conclusions we draw from our samples suspect. Fortunately, there is a solution to this problem that is both technically rigorous and satisfying to common sense: replication. Replication is repeating the assessment again - using significantly different samples of the population. For example, if this year's assessment finds that 100% of student responses meet or exceed our expectations, we do not simply congratulate our students and ourselves and move on. We measure again - perhaps next year - to see what we find. If this remeasure (the replication) finds an equally satisfactory result, we have substantial evidence that students are in fact consistently achieving the desired level of performance.

In practical terms, the recommended procedure* is:

1. collect a representative sample of the sample frame
2. ensure that the size of the sample is adequate
3. replicate

*SACs may choose to use more rigorous standards than those identified here; alternative standards that are routine in their discipline; or alternatives that are derived from their SAC's expertise. In all cases, if SAC's are using the recommendations here or more rigorous ones, elaborative explanations are not required in this report.
Use the following definitions to select the appropriate box to check in this section:

**Random Sample**: each member of the sampling frame has an equal chance of being included in the sample. The textbook example of random sampling is ‘pulling names out of a hat’. Random sampling is the ‘gold standard’ of sampling and should be used whenever possible. The easiest way to generate a random sample is to assign a number to each member of your sample frame and then to use a random number generator to select the sample. ([http://www.random.org/integers/](http://www.random.org/integers/)) Note - to do this, you'll need to know the sample size (see item I [Help Link 4] immediately below). With large sample frames, this can sometimes be difficult.

**Systematic Sample**: this type of sample has two essential features: 1) start with a particular member of the sample frame; 2) select each 'nth' member of the sample frame until the desired sample size has been achieved. These sorts of samples are often considered 'as representative' as random samples. In many/most cases they probably are. The qualifications are crucial: these sorts of samples probably are as representative - but there is really no way to assess whether any bias has been introduced. That being said, bias probably has not been introduced..... Systematic sampling is a recommended sampling strategy for our assessment projects.

An easy way to generate a systematic sample of student data is: 1) determine the desired sample size (see item I [Help Link 4] below); 2) assuming, e.g., that data from multiple student sections will be used, identify the sections and determine how many samples from each section are required; 3) for each section, select a random starting point using the random number generator ([http://www.random.org/integers/](http://www.random.org/integers/)); once the starting point is selected, sequentially select each nth member of the section for the sample. For example; section 54321 has 31 students enrolled when the sample is taken. Since each section is to contribute 5 samples to the project, the course instructor needs to systematically select 5 artifacts. She consults the random number generator and starts with the 23rd name on the roster. She then selects 4 more artifacts (every 6th name on the roster): student 29, 4, 10, 16.

**Stratified Sample**: when a SAC wants to ensure that important, pre-identified groups are proportionally represented (e.g., sex, gender, years of school, previous coursework completed, etc.) they can ensure this by using a stratified sample. First, divide the sampling frame into non-overlapping sub-groups. Once this has been done, continue with either random or systematic sampling as described above. This type of sampling likely will take significantly more organizational effort than the previous two.

**Cluster Sample**: cluster sampling is characterized by dividing the sampling frame into naturally occurring, non-overlapping clusters. In most cases at PCC, clusters might be individual sections of a course. Once the clusters are identified, a random sample of the clusters is selected and all members of the particular cluster are included in the sample. For example, 3 sections of a course are being offered when the student sample is to be drawn. The recommended sample size is 25 (see item 4 immediately below). Each section has at least 25 students enrolled. One of the 3 sections is randomly selected (using, e.g., the random number generator). Once the section is selected, all students enrolled in it will be used in the sample. The bias-introducing problem with this sort of sampling is that unique features of this section (face-to-face or web; instructor reputation; unique instructor techniques; unique textbook; student differences based on when the section is offered [evening class]; etc.) will not accurately represent the whole sampling frame.

**Voluntary Response Sample**: this sampling strategy consists of using self-selected volunteers to participate in the assessment. Any characteristics of the volunteers that are not present to the same
A degree in non-volunteers will introduce bias in the sample. Characteristics such as strong feelings; conscientiousness, desperation for extra credit points; etc. can lead to inaccurate generalizations from these sorts of samples. Sometimes, when assessors are unsure of what the issues might be, the use of voluntary response samples can be helpful in determining what questions might be more rigorously assessed in subsequent projects.

**Opportunity/Convenience Sample**: this strategy consists of using a sample that is easy to get. If your SAC only has one-or-two faculty actively engaged in the current assessment project and they only use artifacts from their sections (and ignore artifacts from others,) this type of sampling is being used. In cases like this, the project can give information about that/those specific section(s), but, it is not possible to generalize to the larger sampling/frame or population.

3. If you are using one of the bias-introducing strategies (in red), share your circumstances/rationale for this decision in the text entry box at the bottom of this section.

4. Describe your sampling procedure in just enough detail to make your step-by-step procedures clear to a reader.

5. Estimating the size of the sample frame/population is essential for proper planning of the project. SACs may use any strategy for this estimate that suits their purpose. In many cases, a simple way to do this is to: 1) determine how many sections will be offered during the term sampling will occur; 2) estimate the average size of each section at the time the sample will be drawn (by referring to, e.g., last year's enrollment data); and 3) generate an estimate of the size of the sample frame (# of sections) x (average # of students in section) = sample frame size. Other strategies are acceptable, e.g., SACs may reference data from recent years' enrollments in formulating their plan.

An appropriate sample size can be calculated after the size of the sample frame/population has been determined. What counts as appropriate varies based on a number of technical considerations: e.g., the type/number of calculations/comparisons that will be utilized in the assessment and the amount and type of variability in the artifacts can be very important. Beyond these sorts of technical complications, expert opinions vary. In keeping with the LAC philosophy of supporting SAC-determined standards of practice, the recommendations that follow represent the most liberal, flexible standard. Any SAC that utilizes these liberal - though commonly recommended standards - can merely report numbers and indicate that the standards are being followed. SACs are free to use their professional judgment in utilizing other standards. If another standard is used, and that standard is less rigorous than this, please share your SACs rationale in the text box at the bottom of this section. If your SAC opts for a more rigorous set of standards, there is no need to elaborate.

To begin, there are two common approaches to assessment that require two separate ways to estimate sample size.

**Generalizing to a population**: sometimes, SACs are interested in population performance: e.g., can all students who complete course XXX demonstrate an appropriate level of proficiency on outcome y? In projects of this sort, the SAC task is to identify how many student artifacts need to be sampled in order to accurately generalize to the sample frame/population. The easiest way to do this is to use a sample size calculator. The web site provides a very brief introduction to the meaning and importance of each of these variables. If you need additional help understanding these issues, begin by consulting your LAC coach.
When the purpose of your sample size estimate is to generalize to the population/sample frame, use the following* (illustrated below):

1. margin of error \(\leq 10\%\)
2. confidence interval \(\geq 90\%\)
3. sample frame estimate (see item 4 above)
4. response distribution = 50% (in this example, based on previously gathered data, the Psy faculty knew that the expected response distribution is 80%, so that number is used here). Since 50% gives the largest sample size, when details about response distribution are not known, use 50%.

*The standard established here is the most relaxed, common standard available. Your SAC may prefer a more rigorous one. If so, the more rigorous, common standard is a 5% margin of error and a 95% confidence interval. In some cases, this standard can result in substantially larger sample sizes.
Sample Size Recommendations with Sub-Group Comparisons

Making sub-group comparisons: sometimes SACs are interested in making group comparisons in their assessment projects. For example, are distance-learning and face-to-face students attaining similar and acceptable levels of outcome attainment? Does the use of one textbook result in better outcome attainment than the use of another? Recommended group sizes for these sorts of comparisons vary based on a number of technical considerations (e.g., type of statistic used; number of comparisons; type of data; actual size of group differences, amount of variation in performance; etc). For general purposes using common statistics, use the following chart to determine reasonable sample sizes.

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Size of Each Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>30</td>
</tr>
<tr>
<td>Meets Most Standards</td>
<td>50</td>
</tr>
<tr>
<td>Optimal</td>
<td>67</td>
</tr>
</tbody>
</table>

Section 3: Project Mechanics - Inter-rater Reliability

Which method of ensuring consistent scoring (inter-rater reliability) will your SAC use for this project?

- Agreement – the percentage of raters giving each artifact the same/similar score in a norming session
  If you are using agreement, describe your plan for conducting the “norming” or “calibrating” session:

- Consensus - all raters score all artifacts and reach agreement on each score

Though rarely used at PCC, some SACs might occasionally use the consistency measure for determining the similarity of their ratings. Consistency is generally only recommended when measuring student improvement – not for showing outcome attainment (which explains its rarity). See the Help Guide for more information. Check here if you will be using consistency calculations in this assessment.

- Consistency* – raters’ scores are correlated: this captures relative standing of the performance ratings - but not precise agreement – and then briefly describe your plan:
1. When using a rubric, it is important to use multiple, expert raters. Multiple raters are essential to avoid potential measurement error. When multiple raters are used, efforts must be made to ensure that scoring is consistent across them. There are 3 primary ways to ensure consistent scoring - aka inter-rater reliability.

**Agreement** - is the goal when attempting to determine whether a clear, objective standard is being met (for example, attaining a proficiency level of 3 in a core outcome). Use agreement when exact agreement between raters is a reasonable expectation. Agreement reliability:

- enables single scoring of artifacts (each rater does not have to rate each artifact) when there is an established, satisfactory level of agreement (done through a norming process);
- when there is an inability to reach adequate levels of agreement there are most likely rubric problems or needs for further training of raters;
- is an appropriate statistic with nominal (dichotomous scales) and ordinal scales (scales are ordered [e.g., better to worse] but scales do not reveal precise degrees of difference). Most rubrics used in educational assessment utilize these scales.

Percent agreement is the statistic most often calculated to quantify this type of inter-rater reliability. This statistic is often calculated by hand rather than via an excel spreadsheet formula. With only two raters, percent agreement is calculated by totaling the number of items scored identically by each rater and dividing by the total number of items. With three or more raters, percent agreement is calculated by first pairing each rater's score of an artifact with all other raters' scores; then dividing the total number of identical paired ratings by the total number of paired ratings. Contact Michele Marden or your LAC coach if you have questions about how to norm and/or calculate inter-rater reliability.

Not surprisingly, the easiest and most common method of calculating inter-rater agreement can produce misleading results. If your SAC wants to ensure the most accurate test of inter-rater reliability, contact Michele Marden or your LAC coach for help.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>When to Use</th>
<th>Standard</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Agreement</td>
<td>With nominal or ordinal rating scales</td>
<td>75-90% agreement (recommendations vary — higher standard is better)</td>
<td>Familiar</td>
<td>• Calculations are required for each scale item</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Consensus can require extended norming</td>
</tr>
</tbody>
</table>
**Consensus** - when there are a manageable number of artifacts to score, each rater can score each artifact. After initial scoring, through a process of discussion, all agree on the score to assign the student's work. This type of inter-rater reliability is often not practical in educational assessment.

**Consistency** - this type of reliability captures relative - but not specific - agreement among raters. It is appropriate when assessing artifacts for improvement, but not when assessing artifacts for outcome attainment. For example, consistency can be a check of ratings when a SAC is trying to determine which textbook contributes more to learning; but, not when that SAC is trying to assess how effectively students are demonstrating attainment of essential outcomes. As a result, most SACs will rarely, if ever, use this type of inter-rater reliability. Consistency measures capture the degree of relative similarity in ratings, without requiring that ratings be identical. For example one rater may be harder than another and give consistently lower scores, but, if both raters agree on which artifacts are better/worse, they may have high consistency. This type of reliability cannot be calculated without performing some statistical analyses. If you are not sure how to proceed with this, consult with your LAC coach.

### Section 3: Performance Benchmarks

1. Have performance benchmarks been specified?
   - Yes (determined by faculty consensus – all instructors who currently teach the course)
   - Yes (determined by only some of the instructors who currently teach the course)
   - Yes (determined by alignment with an external standard (e.g., standards published by the discipline’s professional organization))
   - Yes (determined by post-requisite course expectations within PCC)
   - Yes (determined by post-requisite course expectations for transfer institution)
   - Yes (other). Describe briefly: 
   - No

   If yes, briefly describe your performance benchmarks, being as specific as possible (if needed, attach as an appendix):

   If no, what is the purpose of this assessment (for example, this assessment will provide information that will lead to developing

   benchmarks in the future; or, this assessment will lead to areas for more detailed study; etc.)?


A benchmark is the standard used for comparing, judging, and evaluating the student performance or product (interpreted broadly - often referred to as artifacts). At PCC, benchmarks are determined by each SAC, as the faculty are the content experts and they are in the best position to establish the standards their students should meet. An essential element of benchmarking for core outcome assessment is identifying the appropriate or expected standard of student performance. Establishing appropriate standards can be difficult. At PCC, each SAC is responsible for establishing the expected standards for their students. In higher education assessment, the usual benchmark level is the "just barely passing" student. Many faculty find this standard concerning at first, but, remember, the benchmark is the 'expected level' rather than the 'optimal level'. As an example, for objective/machine-scored artifacts (like multiple-choice tests), the expected standard might be 100% of items correct (or 80% - or 70%). Artifacts that cannot be machine scored most often require complex evaluations of relative quality (like excellent - satisfactory/competent - developing/needs improving - minimal/absent - etc.). When relative evaluations are required, rubrics (identified performance criteria with scaled ratings) are most often required. If a rubric is being used, include it as an appendix to the plan. If your SAC has established benchmarks, (1) select the appropriate check box(es) and then (2) share it in the text entry box. (3) If benchmarks have not been established, share your SAC’s rationale for using this method of core outcome assessment.

There are recommended ways to establish the reasonable, SAC-wide expectations of student outcome attainment (performance benchmarks). Contact Michele Marden or your LAC coach for assistance in this process.

**Section 3. Project Mechanics (Anonymity, Sub-groups & Scorers)**

<table>
<thead>
<tr>
<th>3C. The purpose of this assessment is to have SAC-wide evaluation of student work, not to evaluate a particular instructor or student. Before evaluation, remove identifying student information (and, when possible remove instructor identifying information). If the SAC wishes to return instructor-specific results, see the Help Guide for suggestions on how to code and collate. Please share your process for ensuring that all identifying information has been removed.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>3D. Will you be coding your data/artifacts in order to compare student sub-groups?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>student's total earned hours</td>
<td>previous coursework completed</td>
<td>ethnicity</td>
</tr>
</tbody>
</table>

Briefly describe your coding plan and rationale (and if you selected 'other', identify the sub-groups you will be coding for).

<table>
<thead>
<tr>
<th>3E. Ideally, student work is evaluated by both full-time and adjunct faculty, even if students being assessed are taught by only full-time and/or adjunct faculty. Further, more than one rater is needed to ensure inter-rater reliability. If you feel only one rater is feasible for your SAC, please consult with an LAC coach prior to submitting your plan/conducting your assessment. Other groups may be appropriate depending on the assessment. Check all that apply.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCC Adjunct Faculty within the program/discipline</td>
</tr>
<tr>
<td>PCC FT Faculty within the program/discipline</td>
</tr>
<tr>
<td>PCC Faculty outside the program/discipline</td>
</tr>
<tr>
<td>Program Advisory Board Members</td>
</tr>
<tr>
<td>Non-PCC Faculty</td>
</tr>
<tr>
<td>External Supervisors</td>
</tr>
<tr>
<td>Other:</td>
</tr>
</tbody>
</table>
1. Student identifying information should always be removed before artifacts are submitted to raters/scorers. SACs should also be diligent to remove instructor-specific information whenever possible.

2. Sub-group analysis can be very useful in assessment projects. Students' previous college experience or coursework can sometimes be central issues in exploring student performance. When a SAC determines that student sub-groups might be relevant, they are encouraged to include those elements in their analysis. Sub-group coding is usually done by 1) identifying sub-groups of interest; 2) assigning numerical place-holders for group membership (e.g., 1 for less than 45 hours earned/2 for 45 or more hours earned); 3) comparing relative group performance. If you are coding your data this way, share your rationale for exploring student sub-groups.

3. Check all boxes that apply.