INTRODUCTION
When the SLOAR is generated at the end of each semester the results come back in a series of bar graphs. Each of these graphs represents the percentages at which students achieved various levels of performance on the assessed course outcomes for a given course. Upon seeing the graphs, people often think about or ask a number of questions:

- What does a “good” result look like on the graphs?
- What does a “bad” result look like?
- What do these graphs really mean?

The unfortunate reality is that there are no definite answers to any of those questions that can be applied generally to all assessments. Instead, the answer to each of those questions depends on the specific assessment in question; only the people who designed, deployed, and applied the assessment can answer those questions with any real clarity. This guide is intended to help you answer those questions as they pertain to your department’s assessments.

HOW TO USE THIS GUIDE*:

- **Section 1: Three Factors that Influence Assessment** – This section breaks down the three factors that have the potential to affect your department’s assessment data. It is important that you read this section of the guide because the Section 2 discussion about interpreting SLOAR graphs is framed according to these three factors. The three factors are:
  - Teaching Methods
  - Design of the Assessment Instrument
  - Application of the Assessment Instrument

- **Section 2: Interpreting SLOAR Graphs** – SLOAR data is presented as a series of bar graphs. This section is intended to help you understand how to look at the data distributions on those graphs and draw conclusions based on them. Five types of data distribution are discussed:
  - The Normal Distribution
  - Distribution Skews High
  - Distribution Skews Low
  - Bimodal Distribution
  - Uniform Distribution

*Click on any of the section or sub-section names to jump to that part of the guide.
THREE FACTORS THAT INFLUENCE ASSESSMENT DATA
When it comes to the data that is generated by an assessment, there are generally three factors that help to determine what that data looks like:

1. The teaching methods used by the instructors for the classes in the assessment
2. The design of the assessment instrument – i.e., rubric or multiple-choice exam
3. The application of the assessment instrument – i.e., how did the assessors/instructors interpret and apply the rubric (this factor should not be an issue for multiple-choice assessments)

TEACHING METHODS
This factor is perhaps the most difficult to quantify precisely but it has the potential to cause quite a bit of variance in your results, especially in courses with high enrollment and lots of instructors. The more people that are involved in an assessment, the higher are the chances that the results will be impacted in some negative way. For example, if there are ten sections of a course being assessed and one instructor teaches three full sections of that course, your results could be skewed if that instructor ignores or does a particularly poor job teaching one of the course outcomes.

The bad news is that it is basically impossible to equalize teaching methods across all sections of a course when there are several instructors involved. The good news is that this provides an opportunity for faculty members to engage in conversations about teaching methods and the course material. For the sake of assessment it is important that all the instructors for a given course be on the same page with regard to subject matter and content. That does not mean that all the instructors have to teach the course in exactly the same way, but it does mean that everyone needs to understand that covering all of the course outcomes is mandatory.

DESIGN OF THE ASSESSMENT INSTRUMENT
In short, there are good rubrics and there are bad rubrics. There are good multiple-choice exams and there are bad ones. The quality of the instrument has an enormous impact on the data generated through it and the literature is very clear on this point – there is no such thing as a perfect rubric or a perfect multiple-choice exam. The process of creating, evaluating, and tweaking assessment instruments is iterative and unending. That said, this is the one factor that your department has the most control over because the same instruments must be used across all sections of a course being assessed.

The Guidelines – Creating Rubrics covers the basics of rubric construction and the Guidelines for Creating Multiple-Choice Assessments covers the basics of multiple-choice instrument creation (both documents are available on the MyPGCC portal by clicking College Resources > Resources for Faculty > Academic Affairs Assessment Committee > Guidelines, or Promising Practices), so those documents are a good place to start when it comes to thinking about potential problems with your assessment instrumentation. However, departments that are interested in taking their assessments to the next level of quality should look at Guidelines for Improving the Quality
Guidelines – Making Sense Out of the SLOAR Data

of Rubric-Based Assessments and Guidelines for Improving the Quality of Multiple-Choice Assessments for more guidance in this area.

APPLICATION OF THE ASSESSMENT INSTRUMENT
In a perfect world every written assessment would be evaluated by a team of objective outside assessors who were trained in applying the rubric being used for that assessment. That type of assessment system is preferred because it helps to increase the reliability of the assessment by maximizing objectivity and consistency in scoring. Unfortunately, in practice assessment scoring of that type rarely happens because it is time-consuming and expensive. Instead we have to reply upon a system in which numerous instructors evaluate the quality of their students’ work according to their own understanding of a rubric. Without some deliberate effort to bring consistency to the assessment evaluation process, you run the risk of ending up with data that is skewed in some way.

The solution to this problem of consistency and reliability is having all of the instructors involved in an assessment be on the same page when it comes to how the rubric should be applied. Just like with teaching methods, this requires that faculty members in a given department have conversations about what the rubric means and how it should be applied. Refer to Guidelines for Improving the Quality of Rubric-Based Assessments for some tips on how to do that.
Guidelines – Making Sense
Out of the SLOAR Data

INTERPRETING SLOAR GRAPHS
When your department receives SLOAR data it will be presented as a series of graphs. The distribution of data on those graphs could tell you a lot about how a particular course assessment was completed. However, it is important to note that it is somewhat rare to find a graph that looks exactly like one of the types described here; usually you will get a graph that looks similar to one of these types. It is up to you to determine which of the following examples is closest to the data you are reviewing. Five possible types of data distribution are discussed here.

THE NORMAL DISTRIBUTION
What is it?
In a normal distribution, the Average performance level will be populated by the highest number of students while the Excellent and/or Unsatisfactory levels will be populated by the fewest. That type of distribution is characteristic of data that is normally distributed and it could be evidence that the assessment was a good one. In theory this is the gold standard when it comes to the distribution of scores across a population. If you were to administer an assessment to all of the students at PGCC – all instructed using the same methods and evaluated according to an outstanding rubric that was applied with perfect consistency – you should end up with a distribution of scores that looks something like Figure 1. Still, there are a lot of factors that could prevent your data from being normally distributed.

Figure 1
An Example of a Normal Distribution

What does it mean?
You might say “That graph has lots of students scoring at the Average level. Don’t we want more of our students to score at the Excellent or Good levels?” In general, yes. However, theories of statistical probability tell us that the majority of a given population should end up somewhere near the middle when it comes to a particular question. Think of height and intelligence – the majority of the population would consistently end up near the middle of a
distribution framed by those two variables. We expect no different from assessment data, with certain exceptions.

The most common factor preventing the normal distribution of data in an assessment is probably low sample size. For instance, if your department assessed a course with only one section and 15 students, we would not necessarily expect the data to be distributed normally; there just aren’t enough students in the sample for it to reach normality. However, if you have a large enough sample size (usually \( N \geq 30 \)) and still are not seeing something akin to a normal distribution, it is possible that something in your teaching methods, assessment instrument, or assessment application, or all three, is causing the data to skew high or low.

To be fair, not all assessments, or parts of those assessments, will or should generate data that is normally distributed. For example, if there is a domain in your rubric that calls for students to cite their sources in a particular format, we would hope and expect that a high number of students would achieve a performance level of Excellent on that domain. On the other hand, multiple-choice assessments should generate data that is close to normally distributed; if they do not the assessment might be too easy or hard.

**What should you do?**

In general you want to see SLOAR graphs that are normally distributed. That said, very few SLOAR graphs will show a perfectly normal distribution. If you get a graph that shows something close to a normal distribution, there are still some questions you could ask that could reveal some potential issues with the assessment being analyzed.

1. **Is a normal distribution appropriate in this case?** For example, if you are looking at a rubric-based assessment and there is a course outcome related to the citation of sources, do you really want to see a normal distribution there? Instead, this might be a case where we would prefer to see a data distribution that skews high (see the section named Distribution Skews High for more information). A normal distribution on a course outcome like that might be an indication that the instructors for the course should adjust their teaching methods to spend more time in class discussing the use of proper source citations.

2. **Did all of the instructors who entered data on the assessment turn in data that is somewhat normally distributed?** Department chairs can generate reports in Tk20 that will show the rubric data for individual sections of a course. The report that will yield this data is *PGCC Courses 008 Aggregate Report on Observations*. What you want to see is that all of the instructors for an assessed course entered data that is close to normally distributed. However, if some instructors entered data that was normal while others entered data that skews high or low, it might be a situation where the rubric instrument is not being applied consistently by all instructors. This would be a case where it might be beneficial to spend some time in a department meeting talking about how to apply the rubric instrument more consistently (see *Guidelines for Improving the Quality of Rubric-Based Assessments* for tips on how to do so).
DISTRIBUTION SKews HIGH

What is it?
As the graph in Figure 2 indicates, a distribution that skews high occurs when high numbers of students scored at the Excellent and Good performance levels while fewer scored at the Average level and fewer still scored at Below Average and Unsatisfactory. As discussed in the section on normal distributions, there may be some circumstances when it is understandable that data from an assessment skews heavily towards the upper performance levels (e.g., if the course outcome has to do with citing sources). Nevertheless, when it comes to the assessment of course outcomes a SLOAR graph that skews towards the high end of student performance could be evidence that there is a problem with the assessment.

Figure 2
An Example of a Distribution that Skews High

What does it mean?
While it might be tempting to conclude that a graph that appears this way is indicative of a course that was taught by teachers using the same highly effective methods and evaluated consistently and objectively using a near-perfect rubric, that is not likely to be the case; an outcome like that is theoretically possible, but statistically unlikely. In light of the previously mentioned difficulties involved in accounting for different teaching methods, creating a good rubric and/or multiple-choice exam, and applying those instruments consistently across all sections of a course, the more likely outcome is that something in the assessment process is causing the data to skew.

What should you do?
In cases where the SLOAR graph skews high, the simplest conclusion to draw is that either the assessment instrument is too easy or that instructors are being too generous in their application of the rubric. Whichever it is, data of this sort creates an opportunity for departments to tweak their assessments towards a higher level of quality. While one might think that graphs that skew high are the goal in assessment, the truth is that graphs of this
nature are taken more as evidence that there are flaws in the assessment. Here are some steps your department might take if your data skews high:

1. **Is skewed data appropriate in this case? What is your sample size?** As indicated above, there might be some circumstances where it is appropriate that the data skews high (e.g., an outcome related to the citation of sources). Also, what is the sample size from which this data was generated? If the sample size is low \((N \leq 30)\) then there is a high likelihood that the data will be skewed in some way.

2. **Department chairs and DATs must talk to faculty about the assessment.** The department needs to have a conversation with faculty members about how they applied the assessment and the degree to which they found it to be an appropriate assessment tool. For a detailed discussion about how to go about doing that, please refer to *Guidelines for Improving the Quality of Rubric-Based Assessments* and/or *Guidelines for Improving the Quality of Multiple-Choice Assessments*. Both are available on the S: drive and on the Academic Affairs Assessment page on *MyPGCC*.

3. **If you suspect that the assessment was too easy:** Simply put, the assessment instrument needs to be made more challenging. That could mean removing some very easy questions from a multiple-choice assessment and replacing them with more challenging questions (step-by-step instructions for conducting that type of analysis may be found in *Guidelines for Improving the Quality of Multiple-Choice Assessments*), or it could mean reworking the assignment prompt and rubric for rubric-based assessments (step-by-step instructions for conducting that type of analysis may be found in *Guidelines for Improving the Quality of Rubric-Based Assessments*). It does not necessarily mean that the department needs to completely rewrite those assessments, but some revision is likely called for in this case.

4. **If you suspect that the assessment was applied too generously:** It might be difficult to get faculty to admit to this problem, but it is potentially a significant one. There are lots of reasons why a faculty member might give their students consistently high marks, but one of them is that they fear the assessment data will be used against them in some way. In that case it might be helpful to assure faculty members that it is not the policy of PGCC to use assessment data to evaluate individual faculty members. In fact, it will be viewed as a more significant problem if your department’s assessment data consistently skews high.
DISTRIBUTION SKEWS LOW

What is it?
Just like we do not want data that skews high, neither do we want data that skews low. In this case you will see that high percentages of students were scored at the Below Average and Unsatisfactory performance levels. Obviously no one wants to see results like this and it is relatively rare when it happens.

Figure 3
Example of a Distribution that Skews Low

What does it mean?
Since it is not very likely that all of the instructors for the course being assessed in this case were ineffective or that the majority of students were incapable of attaining the course outcomes, the simplest conclusion to draw is that the assessment instrument is too challenging. Whether it is a rubric or a multiple-choice assessment, departments looking at data of this type should consider adjusting the rigor of the assessment.

There is another possible explanation for data of this type as well. If a course with four outcomes was assessed through one assessment—whether rubric-based or multiple-choice—and data for three out of the four outcomes look relatively normal while the remaining outcome has data that skews low, then it might be the case that something in the course design is preventing students from attaining that fourth outcome.

What should you do?
Here are some questions to consider if your data skews low:
1. What is your sample size? If the sample size is low (N ≤ 30) then there is a high likelihood that the data will be skewed in some way.
2. Is the assessment too challenging? For multiple-choice assessments, you should conduct a test and item analysis (see Guidelines for Improving the Quality of Multiple-Choice Assessments for step-by-step instructions in how to do so) to determine whether
Guidelines – Making Sense
Out of the SLOAR Data

or not the test itself could be causing the data to skew. For rubric-based assessments, department chairs and DATs should talk to faculty members involved with the assessment to get their thoughts on the rigor of the assessment.

3. **For rubric-based assessments, did faculty members provide the rubric to students and/or discuss how the rubric would be applied in class?** It could be that students performed poorly on the assessment because they were unclear on what was expected of them. While some have argued that keeping the rubric from students can lead to more authentic assessments, the research and literature on this point is consistent and clear: Sharing the rubric with students is an essential part of the assessment process. Further, scholars on this point endorse the practice of using class time to discuss the rubric and answer questions about its application. Talk to your faculty members to ensure that they are providing the rubric to their students and talking to them about it.

4. **Is the data skewing low for all course outcomes, or just one outcome?** As indicated above, if the data appears normal for all course outcomes except one, it might be the case that something in the way the assignment/rubric is structured or in how the course is taught that is causing the data to skew. For example, the required textbook for the course might do a poor job of addressing the course outcome with data that skews low. In this case you might have to consider either changing the required textbook or recommending to faculty members that they adjust their teaching methods to make up for the deficiencies in the textbook. Alternatively, the assignment/rubric might be structured in such a way as to make it very difficult for students to address the outcome in question – this might occur if you have a rubric that is covering a lot of course outcomes. In this case it might be worthwhile to adjust the size and scope of the assignment.
BIMODAL DISTRIBUTION

What is it?
In many ways this is one of the most problematic distributions you might find in your SLOAR data (see Figure 4). It looks like the opposite of a normal distribution in that the highest percentages of student scores are found at the Excellent and Unsatisfactory levels and the lowest percentage is found at the Average level. There are many possible explanations for ending up with data of this type.

What does it mean?
Your department might end up with data that looks like this if there is wide divergence in the teaching methods employed by the instructors involved in the assessment. For instance, if one instructor teaching the assessed course failed to provide any instruction on the course outcome in question (or provided instruction that was less than ideal) it would be logical to assume that the students in that instructor’s class would perform poorly on that aspect of the assessment, thus driving the “Unsatisfactory” figures higher.

There could also be a problem with the assessment instrument. If an outcome was assessed through a multiple-choice exam but there were only three questions on the exam that aligned to that outcome (the minimum required for multiple-choice assessments at PGCC), then students could then only attain performance levels of

- Excellent (they got 3/3 questions on that outcome correct),
- Below Average (2/3 questions correct), or
- Unsatisfactory performance levels (1/3 or 0/3 questions correct)

for that outcome. In such a case, the Good and Average levels would have zero students in them.
Guidelines – Making Sense
Out of the SLOAR Data

With regard to rubrics, a bimodal distribution could be an indication that the description for
Average performance on the rubric is vaguely worded or does not differ significantly from the
Excellent and/or Unsatisfactory levels. You might consider making the distinction between the
performance levels more clear.

Finally, another possible explanation involves how the assessment instrument was applied. If
some instructors applied the rubric harshly and others applied it more generously, then you
might end up with data that looked like this. In that case it becomes crucial for departmental
faculty involved in the assessment to have conversations about how they are applying the
rubric. Alternatively, department chairs could examine the data for individual instructors and
initiate conversations about how to apply the rubric more consistently.

What should you do?
Here are some questions to consider if you are looking at data that is distributed bimodally.

1. What does the data look like per instructor and/or per section? Department chairs can
generate reports in Tk20 that will show the rubric data for individual sections of a
course. The report that will yield this data is PGCC Courses 008 Aggregate Report on
Observations. What you are looking for is to see how different instructors applied the
rubric. In other words, did some instructors score students very highly while others
scored them very low? If that is the case, then you will need to initiate conversations
with your departmental faculty to talk about how the rubric should be applied.
Whenever possible this should involve all of the faculty members who will be
responsible for applying the rubric. For tips on how to do this, see Guidelines for
Improving the Quality of Rubric-Based Assessments.

2. For multiple-choice assessments, how many questions are aligned to the outcome for
which bimodally distributed data was generated? As indicated above, it could be the
case that there are simply not enough questions on the assessment to get more precise
data. If there is a course outcome aligned with only three questions on a multiple-choice
assessment, then your department should strongly consider adding more questions that
align to that outcome. Five questions is better and seven questions is ideal for
generating quality assessment data (that way a student could miss two questions and
still achieve proficiency), but even four questions is better than three from the
standpoint of determining proficiency.

3. How does faculty feel about the rubric itself? Did faculty members who applied the
rubric have a hard time distinguishing between the various performance levels? In other
words, are there clear differences between Good, Average, and Below Average levels?
Department chairs and DATs should elicit feedback from the faculty members who
applied the rubric to check for problems with consistency.
UNIFORM DISTRIBUTION

What is it?
Data that is roughly equal across all of the different performance levels for the course outcome in question. Data of this type makes it hard to draw any definitive conclusions around teaching methods and/or the quality of the assessment instrument without some deeper examination of the underlying data.

What does it mean?
Unfortunately it could mean a lot of things. If you end up with roughly equal percentages across all performance levels in your data it could be an indication that the rubric was inconsistently applied by the instructors evaluating student work. It could also mean that the design of the rubric on which the data is based is flawed; there might not be enough to distinguish between the different performance levels for the criterion in question. For multiple-choice assessments, data of this type might indicate that the questions aligned to the course outcome being assessed are not good questions and need to be revised or replaced (for instructions on how to figure out whether or not that is the case, consult Guidelines for Improving the Quality of Multiple-Choice Assessments).

What should we do?
The most important thing your department needs to do is to spend some time discussing the assessment as a group. It’s very likely that the faculty members who applied the assessment will have some thoughts on why the data turned up the way it did and/or how to revise the assessment to make it better. Other questions to consider:

1. What does the data look like per instructor and/or per section? Department chairs can generate reports in Tk20 that will show the rubric data for individual sections of a course. The report that will yield this data is PGCC Courses 008 Aggregate Report on Observations. What you are looking for is to see how different instructors applied the
Guidelines – Making Sense
Out of the SLOAR Data

rubric. In other words, did some instructors score students very highly while others scored them very low? If that is the case, then you will need to initiate conversations with your departmental faculty to talk about how the rubric should be applied. Whenever possible this should involve all of the faculty members who will be responsible for applying the rubric. For tips on how to do this, see Guidelines for Improving the Quality of Rubric-Based Assessments.

2. How does faculty feel about the rubric itself? Did faculty members who applied the rubric have a hard time distinguishing between the various performance levels? In other words, are there clear differences between Good, Average, and Below Average levels? Department chairs and DATs should elicit feedback from the faculty members who applied the rubric to check for problems with consistency.

3. For multiple-choice assessments, what changes should be made that might improve the quality of the data? Based on a test and item analysis report (see Guidelines for Improving the Quality of Multiple-Choice Assessments for step-by-step instructions in how to do them), you might determine that some changes to the assessment are necessary.