Analysis

Qualitative vs. Quantitative vs. Mixed Methods: Assessment instruments can provide us with two different kinds of data: quantitative data that allows us to make statistical comparisons, such as a score on a Likert scale or a test and qualitative data that captures a broader range of potential responses, such as an open response question on a survey or quiz. Some assessment instruments have the benefit of allowing for both quantitative and qualitative data.

Quantitative data: This kind of data can benefit from having relatively higher reliability, as the data is constrained to specific scores, and the scoring process can allow for relatively higher assessment process output with lower inputs as the scoring can be mechanized or at the very least simplified through tools like rubrics, decreasing the intensity and timeline for producing this kind of data relative to qualitative data.

Qualitative data: A case can be made that qualitative data has the potential for higher validity as the open responses and flexible analysis of those responses can capture data that would be lost in the more restricted responses and/or scoring that produces quantitative data. However, there can be a greater subjective element to analyzing these responses that not only sacrifices reliability but can also sacrifice validity, and involves a much higher human resource assessment process input, for a given amount of output with a higher intensity and longer timeline for producing this kind of data as scorer needs to read through and categorize a wide range of responses.

Mixed methods: The approach of capturing both qualitative and quantitative data using a single assessment instrument offers the potential to benefit from both the higher reliability of quantitative measures along with the potential for high validity from qualitative data. This occurs at a very high human resource cost and a higher intensity and timeline for involvement in the work. While this can result in lower potential assessment process output in the form of data, it is also possible to maximize the quantitative data output by scoring everything quantitatively while mitigating some of the input costs process intensity and time by only conducting the qualitative analysis on a subset of the full sample.

Category of Data	Higher or Lower Validity	Higher or Lower Reliability	High, Low and type of Process Input Need	Intensity of Process: Roles and Timeline	Higher and Lower Process Output
Quantitative	Depends on the instrument and in some cases scorer	Higher	Lower	Lower Intensity and shorter timeline	Higher
Qualitative	Depends heavily on the scorer	Lower	Higher	Higher intensity and longer timeline	Potentially Lower
Mixed	Depends on the instrument and the scorer		Highest	Highest intensity and longest timeline	Potential for Highest

Measuring learning objectives vs. engagement, vs. attitudes, values, dispositions and habits of mind: The data you collect can also vary in terms of what it is measuring about students. We can measure what the students are actually learning, how engaged they are in the learning process, and how they feel about and approach the learning process. There can be some overlap between these categories, and some instruments can capture more than one kind of data in this regard.

Measuring learning objectives: Learning objectives can be assessed in a wide variety of ways from scoring artifacts of student work with rubrics to standardized testing, but the common denominator is that we are trying to evaluate what our students know and are able to do with their knowledge and skills. This data is critical for determining the educational inputs and outputs in our analysis of student learning. The levels of validity and reliability vary widely with the instruments used as do the input needs, intensity, timeline and outputs, but in general it requires a relatively high investment of inputs, intensity and timeline relative to the outputs.

Measuring engagement: Student engagement represents a measure of the extent to which our students are active participants in their own learning. This is a growing area of data as we move from course evaluations and surveys to online course usage data, student enrollment data, and even biometrics. The important distinction about this data is although we can certainly measure level of engagement as both an educational input and output, we can also measure it as part of the educational experience in a particular course or program of study. Most of the ways in which we measure engagement vary in their validity as a function of the instrument used, but have potential for higher reliability as the measures of engagement are less subjective than some of the measures of student learning, and while they may often require greater technological inputs through survey administration, student records mining and biometrics, they seldom require much input, or intensity on the part of students and faculty relative to the potential data output.

Measuring attitudes, values, dispositions, and habits of mind: For some of our questions about teaching and learning, we are particularly interested in how the educational experience may shape our students' perspectives, priorities, behaviors, and interests more than their knowledge and skills. There are a variety of ways to measure these things including from student reflection, behavioral observations, interviews, focus groups, and surveys. These measurements can overlap with measurements of engagement which could be considered dispositional and like measures of engagement can be measured as educational inputs, as a window into the educational experience, and as an educational output. Validity and reliability vary with the instrument used, as do the investments in terms of inputs, intensity, and timeline and the resulting assessment process outputs.

Kind of Data	Educational Input, Output or Experience	Higher or Lower Validity	Higher or Lower Reliability	High, Low and type of Process Input Need	Intensity of Process: Roles and Timeline	Higher and Lower Process Output
Learning Objectives	Educational Input and Output	Depends on the instrument	Variable lower reliability	Variable higher inputs	Higher intensity long time	Variable
Engagement	Input, Experience and Output	Depends on the instrument	Higher reliability	High technology low other	Lower intensity, short time	Higher potential outputs
Attitudes, values, etc.	Input, Experience and Output	Depends on the instrument	Variable lower reliability	Variable inputs	Variable intensity and time	Variable

Direct and indirect measures: It may be more useful to think of the relationship between indirect and direct measures as a continuum rather than a dichotomy. Some types of data tend to come from directly measuring the learning outcome, engagement or attitude directly in a teaching and learning context, while others tend to rely more heavily on our ability to infer the objective, engagement or attitude without measuring it directly. For instance, an assignment in a course may be designed to directly assess a particular student learning objective, but even if there are multiple assignments like that one in the course, as long as the final course grade takes into account other factors such as assignments that assess other objectives and consideration for course participation, student final grades represent a more indirect measure of the student learning. However, if students in that same class are asked to rate their own student learning, that would be arguably an even more indirect measure as there are likely even more variables that contribute to a student's self perception of their learning including their capacity to accurately assess that learning.

Direct measures: In general, direct measures of student learning are considered to have more validity than indirect measures, although that validity is still dependent on the instrument as is the reliability of the measure. Direct measures tend to have higher assessment process inputs than indirect measures, but those inputs can be more financial in cases where an external test is purchased and in human resources when a locally developed test or assignment is used. The intensity and timeline for the process can be less with external tests and much more with locally developed assessments, but in each case the assessment process output can be limited by the resources required.

Indirect measures: Indirect measures while potentially less valid, can have reasonably high reliability as students tend to perform consistently in areas like grades and to be consistent in their perspective on their own knowledge, skills, engagement and disposition. The assessment process inputs tend to be less along with the intensity and timeline with the potential for relatively high outputs. However, for indirect measures like surveys, the outputs can be limited by response rates.

Category of Data	Higher or Lower Validity	Higher or Lower Reliability	High, Low and type of Process Input Need	Intensity of Process: Roles and Timeline	Higher and Lower Process Output
Direct	Higher depending on instrument and sometimes scorer	Depends on the instrument	Higher	Higher Intensity and longer timeline	Lower
Indirect	Depends heavily on the scorer	Potentially higher	Lower	Lower intensity and shorter timeline	Potentially Higher

Customized vs. Standardized: Assessment instruments can differ in terms of whether they provide a standardized set of data that can be compared across multiple administrations of the same instrument at different campuses and over time or whether they are unique to a particular time and place limiting the potential for comparisons. The benefits of an assessment instrument that is unique to a particular time and place is that it is flexible and can be customized to reflect the assessment questions being asked in a particular year at a particular institution. However, while standardized tests are often thought of as exclusively containing close-ended responses to insure comparability, this is not always the case as comparability is also created through the use of rubrics and scorer norming and training. Therefore, the distinction between standardized assessments and other customized assessments is primarily related to this issue of flexibility.

Customized: A customized assessment is one that can be or has been changed to the needs of a particular institution, program or year. Sometimes these are also referred to as home-grown or locally developed assessments. They have the potential to have higher validity because they can be designed to address the specific learning objectives or other measures in which a program is interested. Furthermore, the flexibility of design helps insure they can be made relevant to the student learning experience, so students take them seriously. This flexibility comes at a cost in terms of relatively higher assessment process inputs, particularly in terms of human resources and higher assessment process intensity, particularly through faculty roles, and longer timelines because of the development time needed. Finally, the scoring of these assessments may be harder to automate in any way, potentially reducing the assessment process output.

Standardized: Standardized assessments offer common prompts, questions and tasks from year to year and institution to institution. They tend to undergo rigorous testing to insure validity and reliability and as a result are likely more reliable than more flexible assessments. However as noted above their validity while excellent for the objectives they are designed to assess, may not be so strong for the objectives you actually want to assess. Furthermore, if the standardized assessment is an externally developed instrument, standardized tests may not engage student interest and motivation the way a customized, locally developed or adapted instrument can, further limiting their potential validity. In many cases the assessment process inputs are financially high, but not high in terms of human resources, and the process intensity can be lower and the timeline shorter. As a result they offer the potential for higher assessment process outputs. The potential to make comparisons across institutions is very appealing for the

subsequent analysis of the data, but can also be mitigated if different institutions have different priorities.

Category of Data	Higher or Lower Validity	Higher or Lower Reliability	High, Low and type of Process Input Need	Intensity of Process: Roles and Timeline	Higher and Lower Process Output
Customized	Higher depending on what needs to be assessed	Depends on the instrument	Higher	Higher Intensity and longer timeline	Lower
Standardized	Lower if instrument does not match local objectives	Potentially higher	Lower	Lower intensity and shorter timeline	Potentially Higher