HARTWICK COLLEGE

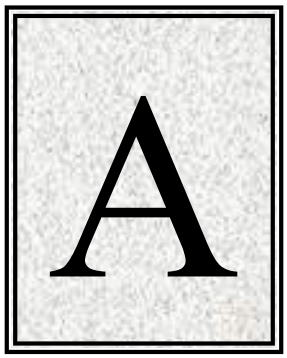
est. 1797

Assessment Plain and Simple: Dictionary of Terms

Like all academic disciplines, assessment has its own language and nuances. The practice of assessment is rooted in the fields of education and the philosophies of social science action research. The following terms are used frequently in assessment. If there is a term that is unfamiliar to you, hopefully this easy, quick-through guide can be useful.

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Looking for a term that is not here? Contact Kimberly Yousey-Elsener (elsenerk@hartwick.edu), Dean of Assessment Retention, for assistance.
Updated Summer 2016



Assessment –an on-going process of:

- Establishing clear and measurable outcomes of student learning or operational effectiveness
- Ensuring that there are sufficient opportunities to achieve those outcomes
- Systematically gathering, analyzing and interpreting evidence to determine how well actual performance matches expectations set forth by the outcomes
- Using the resulting information to understand and improve learning and operational processes (Suskie, 2009, p.4)

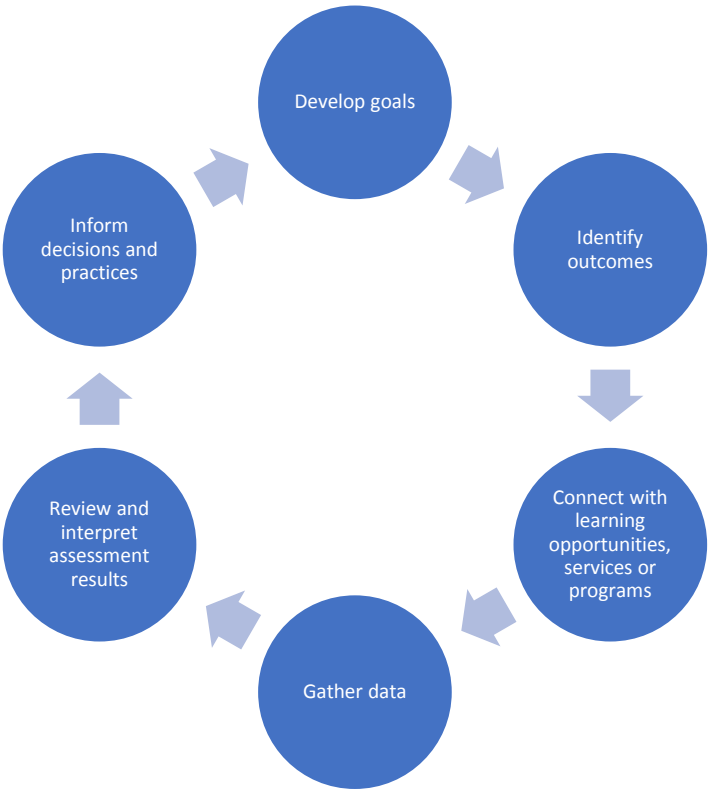
"Good Assessment" - quality assessment typically has the following characteristics:

- Yields reasonable accurate and truthful information so that we can use assessment results with confidence
- Has a clear purpose, so that assessments can be planned and carried out in a focused, simple and cost effective way
- Engages faculty and staff to ask important questions so that results become an important part of decision making
- Flows from clear and focused goals so that results provide useful information that departments and institutions care about (Suskie, 2009, p.37)

Student Learning Assessment - an ongoing systematic process designed by an institution to monitor and improve student learning. It is designed to answer the question: "What will students know, do or feel differently after this experience?" In this process, educators explicitly define what it is they expect students to learn (referred to as learning outcomes); connect those learning outcomes to the learning experience(s) offered; collect data to demonstrate to what extent the learning experience matches the desired outcomes; and use the information to make appropriate modifications in the learning experience (referred to as the feed-back loop).

Institutional Assessment or Institutional Effectiveness - an ongoing systematic process designed by an institution to monitor and improve the quality of the student experience. In this process, departments and programs set goals and gather information that ensures they are meeting stakeholder needs; keeping promises made by the institution's mission and goals; deploying resources effectively, prudently and efficiently (aka being good stewards of resources); serving the public good and/or demonstrating the quality of their work.

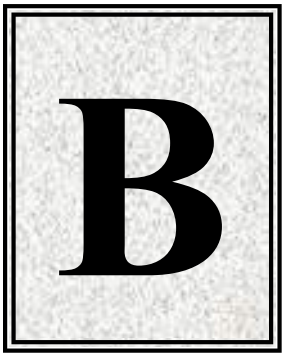
Assessment Cycle - Maps the assessment process as one that begins with developing goals and continues through using data to inform decisions. There are many different assessment cycles, below is a basic example (Chistakis, M. & Bureau, D., 2014, p. 40).



Assessment Plan - a document that maps outcomes with information about how data will be collected; key findings or results of the assessment; and actions for improvement based on the findings.

Assessment tools (Methods) - instruments used to gather data. Tools can be quantitative or qualitative in nature and should be selected in relation to the specific outcome(s) to be assessed; evidence available; learning or process to be measured; and resources available to conduct the assessment. Some familiar assessment tools include: surveys, rubrics, focus groups, tests, essays, and portfolios.

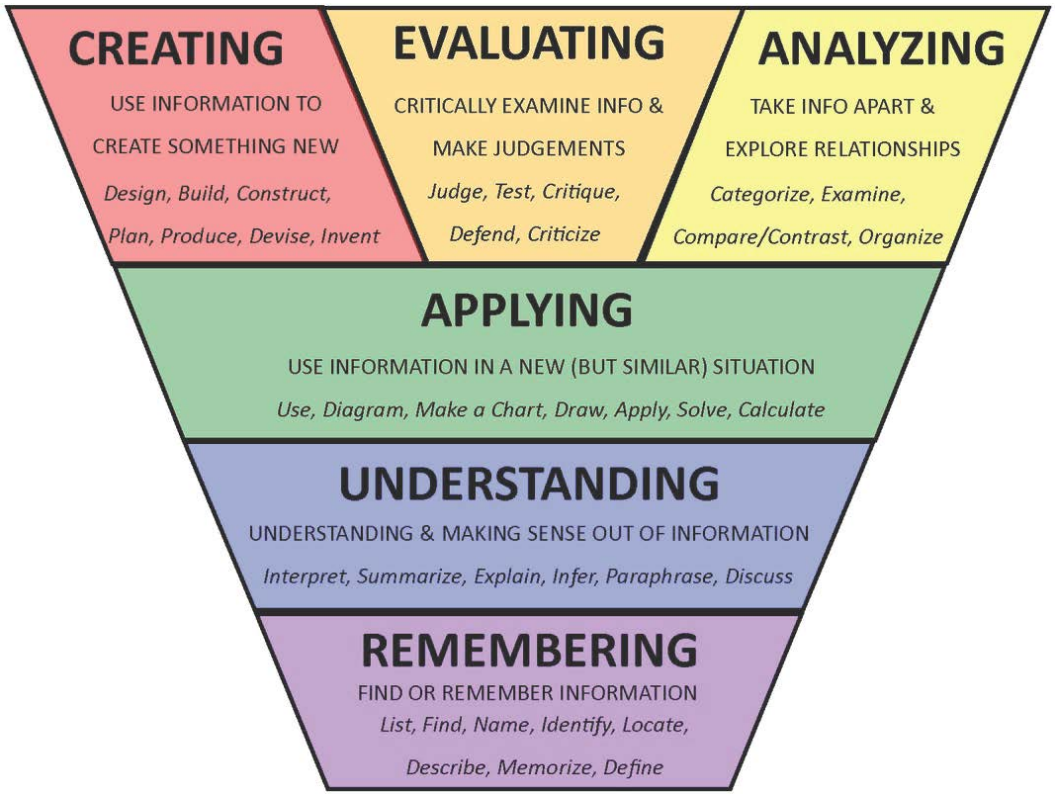
Authentic Assessment - Authentic assessment is a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills. Student performance on a task is typically scored on a rubric to determine how successfully the student has met specific standards (Mueller, J. http://jfmueller.faculty.noctrl.edu/toolbox/whatisit.htm).



Benchmark - In its simplest form, benchmarking is simply comparing data. While many people think of benchmarking solely as large national surveys that are used to compare progress against a national standard or peer group, there are actually many different kinds of benchmark options. These different benchmarks can be grouped into three types:

- **Internal Benchmarking**: Comparing data specific to the institution or department. This data does not contain information from other institutions or organizations, but instead seeks to compare trends within one single organization. Internal data can be tracked over time (e.g., capstone writing rubric scores over the past 5 years) by comparing internal groups (e.g., first year students vs. seniors) or between units at a single institution (e.g., student satisfaction in two different offices with similar functions). The sources for internal benchmarking are typically locally developed and administered surveys and/ or existing institutional data found in tracking systems such as student information systems, judicial records, etc.
- **Best Practices Benchmarking**: Comparing information about common practices in order to learn from others and improve a specific practice or system. This type of benchmark can gather information from other units within an organization, but frequently looks outside to other institutions doing similar work. It is common for best practices benchmarking to be more qualitative or descriptive in nature. Survey data can be used for things like staffing ratios, square footage of facilities, faculty: student ratios, etc., but more often than not best practice benchmarking involves comparing written documentations about programs or services (e.g., websites, marketing materials, etc.) and/or gathering information via interviews with different campuses and individuals.
- **External Benchmarking**: Comparing a given department, unit or institution with national standards, a national average or peer institution group. External benchmarking is usually more quantitative in nature and involves collecting data within an organization and then comparing that same data with an external entity. The sources of data for this type of benchmarking are usually found in national databases such as IPEDS (The Integrated Post-Secondary Education Data System, http://nces.ed.gov/ipeds/) or via commercially developed survey instruments such as the NSSE (National Survey of Student Engagement, http://nsse.iub.edu/). (Yousey-Elsener, 2014, pp. 50-51)

Bloom's Taxonomy- According to Bloom, knowledge increases as one progresses through the levels of learning, where remembering is the most surface level and creating represents the deepest form of learning (see below for description of the levels). The Taxonomy has been considered very useful in defining learning-outcome statements and determining which data collection tool to use. For example, "lower level" outcomes are effectively measured using tools more quantitative in nature, such as tests and surveys vs. "higher level" outcomes that are more effectively measured using qualitative tools such as essays, projects and presentations paired with a rubric.



C

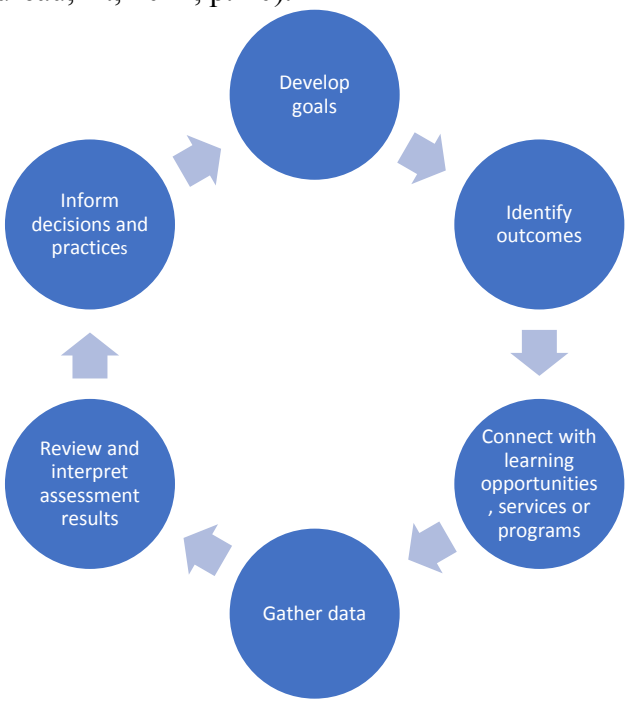
Classroom Assessment - (sometime referred to as Course-based Assessment or Embedded Assessment) - is a process of gathering data on student learning during the educational experience, designed to help the instructor determine which concepts or skills the students are not learning well, so that steps may be taken to improve the students' learning while the course is still in session. This is an example of formative assessment.

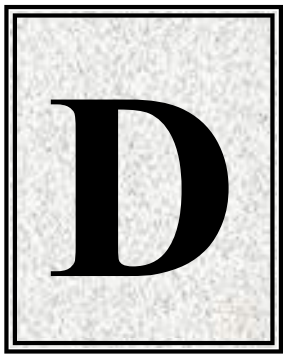
Closing the Loop – A process through which assessment data is shared, reflected upon and put into action through a change in any stage of the assessment cycle (e.g., change in outcome, data collection tool, program/service/learning experience, etc.).

Culture of Assessment – A set of pervasive actions and behaviors by members of the institutional community that focuses on sharing and using data to inform decisions regarding the improvement of programs, services and student learning (Henning, 2015, p. 11). In this environment assessment is a valued and expected part of the decision making process.

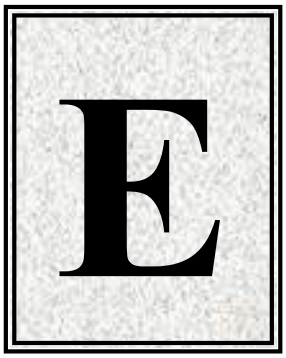
Culture of Evidence – An environment where assessment results are used to guide policy decisions and is expected and valued.

Cycle (or Assessment cycle) – Maps the assessment process as one that begins with developing goals and continues through using data to inform decisions. There are many different assessment cycles; here is a basic example (Chistakis, M. & Bureau, D., 2014, p. 40).

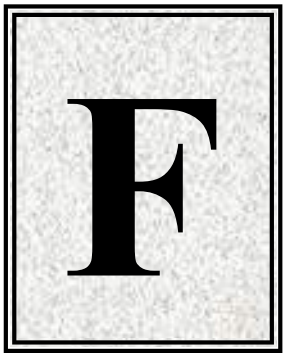




Direct Assessment (opposite of in-direct assessment) –measures that allow you to observe the change that is occurring. In direct measures of learning, students display knowledge or skills as the result of an assessment measure (presentation, test, etc.). Direct measures of student learning require students to display their knowledge and skills as they respond to the instrument itself. Objective tests, essays, presentations, and classroom assignments all meet this criterion (Palomba, C.A., & Banta, T.W., 1999).



Embedded Assessment - (sometime referred to as Course-based Assessment or Classroom Assessment) - is a process of gathering data on student learning during the educational experience, designed to help the instructor determine which concepts or skills the students are not learning well, so that steps may be taken to improve the students' learning while the course is still in session. This is an example of formative assessment.



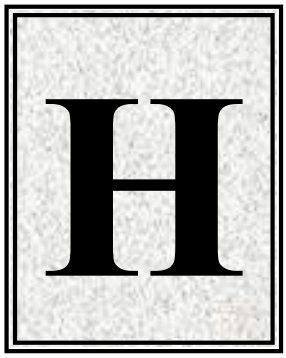
Feedback Loop - use the results of the assessment to make appropriate modifications to a course, program, service or curriculum.

Formative Assessment (opposite of summative assessment) - assessment that occurs during the activity, course or process with the focus on improving as it is occurring.

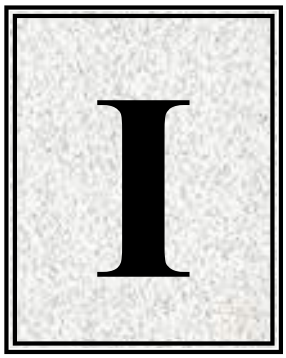
Goals – Goals are general statements about what we need to accomplish to meet our mission or serve our purpose. Goals typically are more broad and longer-term than an outcome. They are often used to help connect outcomes to a larger idea at an institution. There are two frameworks at Hartwick that set goals for the institution, the Organizing Principle and Strategic Framework and the General Education Student Learning Outcomes. All assessment outcomes at Hartwick should be connected with one or both of these Frameworks.

- **Organizing Principle and Strategic Framework** (http://www.hartwick.edu/organizingprinciple)
  - Improve Student Experience and Satisfaction
  - Maximize the Academic Program
  - Expand Our Financial Base
  - Improve the College's Image and Reputation
  - Maximize Employee Performance
  - Maximize College Governance
  - Maximize Financial Performance

- **Hartwick College General Education Student Learning Outcomes: Our Commitment to the Liberal Arts in Practice**. Through their education both within and outside the major, Hartwick graduates will be able to:
  - Communicate effectively in written English
  - Communication effectively in spoken English
  - Communicate in one non-native language
  - Express relationships in formal logical or mathematical language and interpret relationships so expressed
  - Identify and evaluate the consequences of individual and collective values, beliefs, ideas and actions
  - Develop, test and evaluate hypotheses using appropriate information and methods
  - Produce interpretive or problem-solving creative work
  - Apply knowledge through practical experience



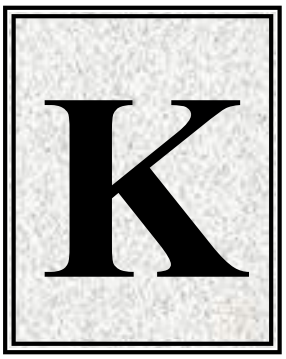
Holistic Assessment - Making judgments about student learning by using an overall appraisal of the students' entire performance, rather than scoring or analyzing components of student performance individually; the complete or final product of the learning experience is assessed as a whole. This is usually done using a portfolio or some other collection of student work.



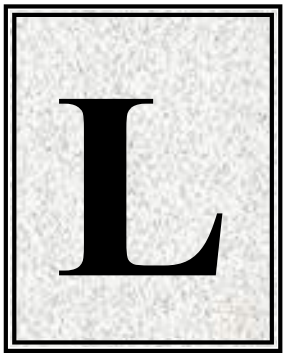
Indirect Assessment (opposite of Direct Assessment) - evidence is inferred instead of being supported by direct evidence (i.e., usage data, satisfaction surveys) (Palomba, C.A., & Banta, T.W. ,1999). Indirect assessment usually focuses on gathering students' opinions, perceptions or reflections on their experience.

Institutional Assessment or Institutional Effectiveness - an ongoing systematic process designed by an institution to monitor and improve the quality of the student experience. In this process, departments and programs set goals and gather information that ensures they are meeting stakeholder needs; keeping promises made by the institution's mission and goals; deploying resources effectively, prudently and efficiently (aka being good stewards of resources); serving the public good and/or demonstrating the quality of their work.

Inter-rater reliability - when two or more raters observe an individual's behavior, record scores, and then the scores of the raters are compared to determine whether they are in agreement (Creswell, J., 2002). Inter-rater reliability is particularly important in situations where a rubric is being used by multiple people, it's important to test the rubric and train the raters to ensure they are consistently interpreting the rubric.

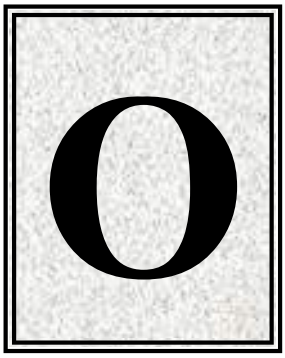


Key performance indicator - (also referred to as metric, milestone or target) Numbers that document how well operational aspects of a program, service or institution are functioning. They often track progress towards larger goals. Common metrics that you find in higher education are: retention rate, graduation rate, enrollment numbers, alumni giving, faculty ratios, etc.



Learning outcome - (sometimes referred to as objectives) - statements that describe specific behaviors a student is expected to demonstrate as a result of a learning experience either in the classroom or outside of the classroom. Learning outcomes are often phrased in what a student should know, think, feel or be able to do (e.g., skills) upon completion of a course, program, or curriculum. Learning outcomes at Hartwick should be linked to the General Education Learning Outcomes:

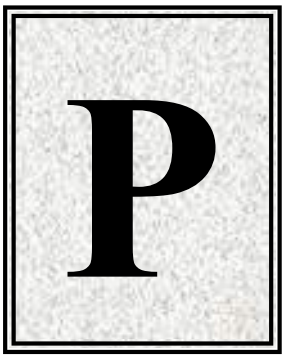
- **Hartwick College General Education Student Learning Outcomes: Our Commitment to the Liberal Arts in Practice**. Through their education both within and outside the major, Hartwick graduates will be able to:
  - Communicate effectively in written English
  - Communication effectively in spoken English
  - Communicate in one non-native language
  - Express relationships in formal logical or mathematical language and interpret relationships so expressed
  - Identify and evaluate the consequences of individual and collective values, beliefs, ideas and actions
  - Develop, test and evaluate hypotheses using appropriate information and methods
  - Produce interpretive or problem-solving creative work
  - Apply knowledge through practical experience



Outcome – specific statements that bound broad goals to a place, time and group of participants. It can sometimes also specify the target or standard that is intended to be reached. They specifically state what you want the result of your efforts to be or the changes you want to occur. There are generally two types of outcomes: **learning and operational**. Learning describes what a student should know, think, feel or be able to do as the result of a learning experience. Operational outcomes describe a change in a process, procedure, office, system or quality of an experience. Strong outcomes are often referred to as S.M.A.R.T. outcomes:

- **S – Specific** – anyone can read your outcome and know exactly what you are hoping to accomplish. Therefore it's important to pay close attention to using phrases that are often used in higher education but many people could disagree on the meaning, for example – critical thinking, leadership, communication skills, learn, understand. If you find your outcome using these words, select a new verb that better defines what that concept would look like in your setting and context.
- **M – Measurable** – it can easily be determined how one might gather data that can determine if what is happening matches the intended outcome. Sometimes measuring an outcome is easy; you can look at the outcome and say we'll do a survey, or create a rubric, etc. Sometimes it takes some thought and getting ideas from colleagues. If you are finding the outcome is not measurable, you may need to refine the outcome to be more measurable.
- **A – Achievable** – the outcome can be accomplished in a specific time and place. If you find your outcome is too big (aka cannot be accomplished given your resources and time period), then it is not achievable. In addition, for learning, this means that the outcome is connected with a specific learning experience. For operational outcomes, this means that the outcome is connected with a specific process.
- **R – Relevant** – put simply this means the outcome is meaningful to the department, program and/or institution. It should have value and people should care about it.
- **T – Time Sensitive** – the outcome is happening within a specific time period, most of the time that is within a semester or academic year. (Yousey-Elsener, 2013, p. 21)

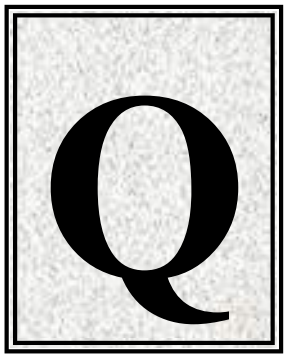
Operational Outcomes (sometimes called Program Outcomes) – statements that describe specific expectations for change in an office, process, system or experience. These outcomes focus on office operations or student experiences that are occurring outside of the learning. They often set expectations around efficiency, effectiveness, climate, needs and/or satisfaction.



Program Outcome (sometimes called Operational Outcomes) – statements that describe specific expectations for change in an office, process, system or experience. These outcomes focus on office operations or student experiences that are occurring outside of the learning. They often set expectations around efficiency, effectiveness, climate, needs and/or satisfaction.

Program Review – Program review is a comprehensive evaluation of an academic program that is designed both to foster improvement and demonstrate accountability. Program reviews typically include a self-study conducted by the program's faculty and staff, a visit by one or more external reviewers, and recommendations for improvement based on the conclusions of the self-study and the reviewer (Suskie, 2009, p. 14).

Practical Significance – When sample sizes are very large (over 1000) or very small (less than 100), data may show a statistical significance with the slightest difference between groups. In this case, it is important to ask a larger question about whether the differences are meaningful in a specific context. Data can be practically significant without being statistically significant as well as the opposite – statistically significant but not practically significant. For example, a campus may find that statistical significance reported on the NSSE is not practically significant because it does not apply to the mission or priorities of the campus (Yousey-Elsener, 2013, p. 144)

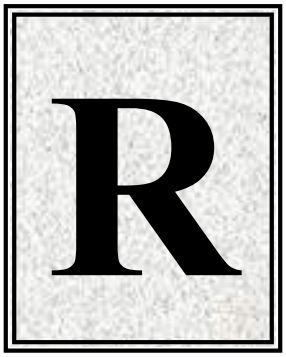


Qualitative tools/data - ways of collecting information that are concerned with understanding or conveying meaning or contexts, rather than making statistical inferences. Qualitative tools tend to focus on collecting data that is word-based and offer a *depth of understanding* of a topic. They are particularly useful when looking at learning outcomes that call for higher-levels of learning* (e.g., creating, evaluating, analyzing, applying) as well as operational outcomes that are focused on understanding why or how something is occurring. Common forms: open-ended survey questions, essays, presentations, portfolios, participant observations focus groups, in-depth interviews, etc.

*See Bloom's Taxonomy

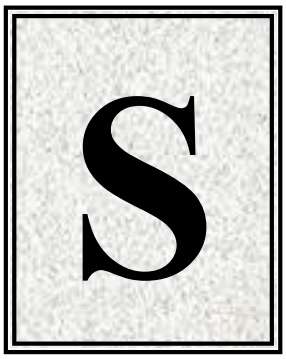
Quantitative tools/data – information that is collected or represented numerically that are concerned with counting occurrences or measuring characteristic's or behavior rather than their meanings. Quantitative tools provide data that is easy to analyze statistically and offer a *breadth of understanding* related to a topic. They are particularly useful when looking at learning outcomes that call for lower-levels of learning* (e.g., remembering and understanding) as well as operational outcomes that are focused on understanding the who, what, where of a topic. Common forms: surveys, tests, quizzes, some rubrics, existing data in information systems, experiments, questionnaires, etc.

*See Bloom's Taxonomy



Reliability – Degree to which an assessment tool measures the same thing each time it is used (Yousey-Elsener, 2013, p. 144). For example, if a rubric tool is used to measure student learning related to critical thinking for the same assignment in multiple sections of a course, the results should be consistent across all sections. Another example might be a survey that is administered every year to all students on campus; each year the survey should yield similar results, with some movement in scores but not larger movements. There is a statistical test one can use to test for reliability in data, however, that test requires a relatively large data set in order to be effective. In assessment we tend to collect smaller data sets so we rely on our own ethical judgement and observations in order to determine reliability.

Response Rate – "The percentage of persons in the sample who completed and returned the survey" (Lodico, Spaulding, Voegtle, 2006, p. 170). For example, if survey is sent to 100 people and 57 people complete the survey, the response rate is 57%. The national average for response rates is currently around 20% and falling. Survey fatigue is most often cited as the reason for declining response rates. Regardless of the number of responses, caution should be used when focusing on response rates. It is always important to look at who responded in relation to the overall group to determine if your respondents reflect the overall sample or group. If your response rate is 75% but the response pool is lopsided with white, female students it may not be an accurate reflection of the overall group.



Self-Assessment - (sometimes referred to as reflective learning) - a process in which a student engages in a systematic review of his or her own performance or learning, usually for the purpose of improving their learning in the future.

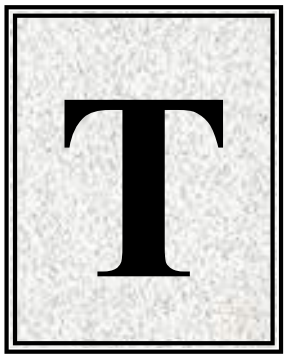
S.M.A.R.T. outcomes – Learning or operational outcomes should meet the following criteria:

- **S – Specific** – anyone can read your outcome and know exactly what you are hoping to accomplish. Therefore, it's important to pay close attention to using phrases that are often used in higher education but many people could disagree on the meaning, for example – critical thinking, leadership, communication skills, learn, understand. If you find your outcome using these words, select a new verb that better defines what that concept would look like in your setting and context.
- **M – Measurable** – it can easily be determined how one might gather data that can determine if what is happening matches the intended outcome. Sometimes measuring an outcome is easy; you can look at the outcome and say we'll do a survey, or create a rubric, etc. Sometimes it takes some thought and getting ideas from colleagues. If you are finding the outcome is not measurable, you may need to refine the outcome to be more measurable.
- **A – Achievable** – the outcome can be accomplished in a specific time and place. If you find your outcome is too big (aka cannot be accomplished given your resources and time period), then it is not achievable. In addition, for learning, this means that the outcome is connected with a specific learning experience. For operational outcomes, this means that the outcome is connected with a specific process.
- **R – Relevant** – put simply this means the outcome is meaningful to the department, program and/or institution. It should have value and people should care about it.
- **T – Time Sensitive** – the outcome is happening within a specific time period, most of the time that is within a semester or academic year. (Yousey-Elsener, 2013, p. 21)

Statistical Significance – Statistical test conducted to determine if the differences in group averages are due to chance or are "actually" occurring. Typically anything with a significance level of .05 or lower (sometimes represented as $p<.05$) is considered statistically significant. In assessment it is also important to consider **practical significance**.

Student Learning Assessment - an ongoing systematic process designed by an institution to monitor and improve student learning. It is designed to answer the question: "What will students know, do or feel differently after this experience?" In this process, educators explicitly define what it is they expect students to learn (referred to as learning outcomes); connect those learning outcomes to the learning experience(s) offered; collect data to demonstrate to what extent the learning experience matches the desired outcomes; and use the information to make appropriate modifications in the learning experience (referred to as the feed-back loop).

Summative Assessment - process for measuring the overall level of student learning or operational effectiveness at the end of an experience or course. Its primary goal is to collect information that can be used to improve future experiences, but does little to affect the experience of the people currently participating. Course evaluations are an excellent example of a summative assessment. They are conducted at the end of the semester, and the data provided to faculty after the course ends in order to make improvements the next time the course is offered.



Transferability – "…the degree of similarity between the…[assessment site, context] and other sites as judged by the reader" (Lodico, Spaulding, Voegtle, 2006, p. 275). Transferability is important when considering best practices or assessment tools used at other campuses or in a different department. Sometimes they can be easily borrowed and transferred over to your context and sometimes they need to be adapted in order to meet your needs.

Trustworthiness – like validity and reliability for quantitative data, trustworthiness concerns itself with how rigorously qualitative data has been collected and analyzed. When analyzing qualitative data, it is important to keep in mind the following:

- **Credibility**: are you representing the data in a manner that is true to the source that provided it? Was the data collected in an ethical manner?
- **Dependability**: Have you collected and analyzed data in a systematic and consistent way that can be demonstrated to others? Did you keep notes about decisions you made when analyzing data?
- **Transferability:** Have you provided enough detailed descriptions that someone can determine if the results can be applied in their context?
- **Promoting action:** Have you used the data to make a positive change? (Yousey-Elsener, 2013, p. 91)

Validity – Does the instrument used measure what it is intended to measure? If you have ever collected data and then said – well that's not what I was looking for! – then you have run into validity issues. Fear not, sometimes that data can be used to inform a different question or outcome, but you will have to take a close look at the instrument if you need data for the original outcome. There are statistical tests that can be run to measure validity, but they require a large data set. Therefore, in assessment we focus on face validity and content validity.

- **Face validity**: "the instrument appears to be measuring what it intends to measure." This means that on the surface, the questions seem to fit whatever is the described purpose of the instrument (Lodico, Spaulding, Voegtle, 2006, p. 111). This can be determined through a pilot, or by asking peers with knowledge of the purpose and instrument design to provide a review of the instrument.
- **Content validity**: "ensures the instrument is measuring the breadth and depth of the issue that it is intended to measure" (Lodico, Spaulding, Voegtle, 2006, p. 112). When conducting research, content validity is usually tested during the pilot process. For assessment, piloting an instrument is sometimes not needed, therefore one can have the instrument reviewed by a peer who has specialized knowledge in the topic to determine if the instrument is leaving out any important questions related to the topic and if questions included on the survey are appropriately related to the topic.

References

Christakis, M. & Bureau, D. (2014). Tenet three: Developing infrastructure for student affairs assessment practices. In K.M. Yousey-Elsener, E.M. Bentrim, and G.W. Henning (Eds.), *Coordinating student affairs divisional assessment: A practical guide* (pp. 11-34). Sterling, VA: Stylus Publishing.

Henning, G.W. (2015). Tenet two: Cultivating a culture of assessment. In K.M. Yousey-Elsener, E.M. Bentrim, and G.W. Henning (Eds.), *Coordinating student affairs divisional assessment: A practical guide* (pp. 11-34). Sterling, VA: Stylus Publishing.

Lodico, M.G., Spaulding, D.T., Voegtle, K.H. (2006). Methods in educational research: From theory to practice. San Francisco: Jossey-Bass.

Palomba, C. A., & Banta, T. W. (1999). Assessment essentials: Planning, implementing, and improving assessment in higher education. San Francisco, CA: Jossey-Bass.

Suskie, L. (2009). Assessing student learning: A common sense guide. San Francisco: Jossey-Bass.

Yousey-Elsener, K.M. (2013). Successful assessment for student affairs: A how-to guide. Little Falls, NJ: PaperClip Communications.

Yousey-Elsener, K.M. (2014). Assessment: Data collection and reporting. Little Falls, NJ: PaperClip Communications.